

VOLUME 20 NUMBER 2 JUNE 1957

---

# SOCIOMETRY

**A JOURNAL OF RESEARCH  
IN SOCIAL PSYCHOLOGY**

## *Articles by*

**FIEDLER  
BORG  
BELL  
FARBER  
COHEN  
WHEELER  
BALES AND SLATER  
GLASER  
SARBIN AND BAKER  
ZILLER**

Published by

**THE AMERICAN SOCIOLOGICAL SOCIETY**

# SOCIOMETRY

## EDITOR

LEONARD S. COTTRELL, JR.

## ASSOCIATE EDITORS

R. FREED BALES	HERBERT HYMAN
HERBERT BLUMER	IRVING L. JANIS
JOHN A. CLAUSEN	FREDERICK MOSTELLER
LEON FESTINGER	THEODORE M. NEWCOMB
NELSON N. FOOTE	WILLIAM H. SEWELL

## EDITORIAL CONSULTANTS

BERNARD M. BASS	PHILIP J. MCCARTHY
EDGAR F. BORGHATTA	RICHARD T. LAPIERE
ORVILLE G. BRIM, JR.	STANLEY SCHACHTER
LAUNOR F. CARTE	HERBERT A. SIMON
FRED E. FIEDLER	SHIRLEY A. STAR
ROBERT N. FORD	FRED L. STRODTBECK
NATHAN L. GAGE	GUY E. SWANSON
HERBERT GOLDFAMER	E. PAUL TORRANCE
LEO A. GOODMAN	JOHN W. M. WHITING
ALEX INKELLES	ROBIN M. WILLIAMS, JR.
WILLIAM W. LAMBERT	ROBERT F. WINCH

## EXECUTIVE OFFICER

MATILDA WHITE RILEY

*Published four times a year*

March, June, September, December

by

THE AMERICAN SOCIOLOGICAL SOCIETY

at The Waverly Press, Mt. Royal and Guilford Aves., Baltimore 2, Md.

*Sociometry was founded in 1937 by J. L. Moreno*

Communications for the Editors and all manuscripts should be addressed to Leonard S. Cottrell, Jr., Russell Sage Foundation, 505 Park Avenue, New York 22, New York. Business communications should be addressed to Matilda White Riley, The American Sociological Society, New York University, Washington Square, New York 3, New York.

Copyright 1957 by THE AMERICAN SOCIOLOGICAL SOCIETY.

Entered at the Post Office at Baltimore, Maryland, as second-class matter.

# SOCIOMETRY

A JOURNAL OF RESEARCH IN SOCIAL PSYCHOLOGY

---

VOLUME 20 • NUMBER 2 • JUNE 1957

## CONTENTS

- A Note on Leadership Theory: The Effect of Social Barriers between Leaders and Followers.....*Fred E. Fiedler* 87
- The Behavior of Emergent and Designated Leaders in Situational Tests.....*Waller R. Borg* 95
- Anomie, Social Isolation, and the Class Structure.....*Wendell Bell* 105
- An Index of Marital Integration.....*Bernard Farber* 117
- Stimulus Conditions as Factors in Social Change.....*Edwin Cohen* 135
- Notes on "Role Differentiation in Small Decision-Making Groups"  
*D. K. Wheeler* 145
- Notes on "Role Differentiation in Small Decision-Making Groups":  
Reply to Dr. Wheeler.....*Robert F. Bales and Philip E. Slater* 152
- A Note on "Differential Mediation of Social Perception as a Correlate of Social Adjustment".....*Daniel Glaser* 156
- Psychological Predisposition and/or Subcultural Participation: Reply to Dr. Glaser.....*Theodore R. Sarbin and Bela O. Baker* 161
- Group Size: A Determinant of the Quality and Stability of Group Decisions.....*Robert C. Ziller* 165

# SOCIOMETRY

## EDITORIAL POLICY

*Sociometry* is concerned with the entire range of interests and problems represented by research in social psychology. It is the policy of the editors to seek those manuscripts for publication which represent the significant research interests of investigators who are concerned with giving the field of social psychology theoretical structure and reporting research which is clearly focused, well designed, and competently conducted.

While social psychology is presently regarded by most as a field with indeterminate boundaries, it has as its central focus the investigation of the processes and products of social interaction at the interpersonal, intrapersonal, intergroup and intragroup levels and the development of significant generalizations therefrom. In keeping with the more general meaning of the name of the journal emphasis will be placed on measurement of social behavior. However, this emphasis does not exclude the acceptability of good articles which must rely upon qualitative materials and analyses.

The editors and editorial consultants can be expected to subject manuscripts to rigorous criticism and screening according to the best standards of scientific research and at the same time avoid a sterile orthodoxy which would stultify the communication of creative ideas at the growing edge of the science. Thus the journal will strive to be flexible in its response to the publication needs of its contributors.

It is the intention of the editors to avoid any tendency toward professional provincialism and to invite contributions from any sector of the scientific community which promise to further the objectives of the journal.

---

Annual subscription rate \$9.00; special rate to members of THE AMERICAN SOCIOLOGICAL SOCIETY \$4.50. Single issues \$2.50.

*Specifications for Manuscript Preparation for Sociometry* may be obtained from the Editorial Office.

*Back Issues* dated 1937 through 1955 should be ordered from the former publisher, Beacon House, Beacon, New York.

Notice of change of subscriber's address should be sent to THE SOCIETY four weeks in advance, indicating the old as well as the new address.



## A Note on Leadership Theory: The Effect of Social Barriers between Leaders and Followers<sup>1</sup>

FRED E. FIEDLER, *University of Illinois*

The military customs which limit social intercourse between leaders and followers are probably as old as organized armies. Officers and enlisted men have "always" had different mess halls, different sleeping quarters, different clubs, and different washrooms. To a somewhat lesser extent, this is also true of senior and junior officers and of higher and lower ranking enlisted personnel. These barriers to informal interaction are considered sufficiently important so that they are observed even on submarines where space is at a premium.

While military organizations provide a good illustration, this probably applies to some extent to all formally structured groups. Thus, restrictions limiting social interaction between executives and employees are present in most large organizations. Many corporations operate executive dining rooms and employee cafeterias; hospitals frequently provide one set of dining and lounge facilities for the medical staff and one for the nursing and lower administrative echelons; and many older university buildings still maintain separate dining and washrooms for faculty and for students.

After the war, the military services made several attempts to implement recommendations by the Doolittle commission to lower social barriers between officers and enlisted men. But these experiments were soon abandoned as not conducive to the maintenance of proper discipline. In light of the long history of these conventions (cf. 6, p. 94; 17, p. 37), as well as some of the unsuccessful attempts to change them, it seems appropriate that we re-evaluate the functions which these customs might serve.

A number of explanations are generally advanced to justify these barriers. One is that separate facilities give greater privileges, hence, greater reward, to men of higher rank. However, this is obviously only part of the story: facilities for higher ranks are frequently, but not always, superior to those for lower ranks; second, the privileges are not permissive. For example, officers are prohibited from drinking or gambling with enlisted men. To say that separate facilities provide greater privacy begs the question, "Why is greater privacy required?"

A second explanation justifies these customs by the adage that "familiarity breeds contempt." In other words, discipline and group productivity

<sup>1</sup> This paper is based on research supported by Contract N6-ori-07135 between the Office of Naval Research and the University of Illinois. The writer is indebted to Dr. Ivan D. Steiner for his many helpful criticisms and suggestions.

can be maintained more readily if superiors and subordinates do not mix socially.

This argument, which will here be examined further, implies that the subordinate will not follow his leader (a) if he has extensive social contacts with him or (b) if he has knowledge of his superior's weaknesses, i.e., if he sees his superior "with his hair down." The knowledge of a superior's weak points is said to lower discipline presumably because the men will be less likely to idealize their leader and they will then refuse to follow him unquestioningly.

By curtailing informal contacts, the subordinate may be prevented from learning something about his superior which he could use for the purpose of pressuring or gently blackmailing him, either to get advancement or special favors.

However, the implication that lack of social contacts would prevent subordinates from knowing their superiors' faults and foibles does not seem very realistic to anyone who has ever worked in an office or shop, or who has served in the armed services. The weaknesses and indiscretions of superiors are surely among the favorite topics of conversation, ranking only slightly below the subject of sex and women. Thus, it is unlikely that the leader's status with his subordinates depends on how much his subordinates know about him, nor is it likely that these rules are designed to protect him from blackmail. While it is said that "no man is a hero to his valet," there have obviously been many faithful and well disciplined valets.

A reconsideration of this problem leads us to propose a contrary view, namely, that the social barriers between superiors and subordinates primarily serve to protect the leader from emotional involvement with his followers. According to this position, a work group will suffer more when the leader becomes too familiar with his men than when the men become too familiar with their leader.

A considerable body of anecdotal evidence supports this notion. Leaders frequently report that they experience conflicts in sending close personal friends on dangerous missions, or that acquaintance with a man's family influences their decisions—both are favorite plots of war films and novels. A close friendship with a subordinate obviously implies the presence of emotional ties which may interfere with the superior's task. The leader will tend to be less capable of dealing in an objective manner with a man with whom he is emotionally involved than with one whom he knows only in line of duty.

#### A REVIEW OF PERTINENT FINDINGS

Recent research also favors the interpretation which we have here advanced. The writer and his associates have conducted field studies on inter-

personal perception and group effectiveness which may throw some light on the problem. Since the studies were not designed specifically with this question in mind, the evidence is not so conclusive as one might wish. Moreover, the studies were conducted over a time span of six years during which the research methodology underwent various changes. For this reason, data collected in one investigation were not necessarily gathered in others. In short, although we grant that the present data may not be sufficiently convincing to settle the question, we hope they will be sufficiently thought provoking to lead to further research.

#### GENERAL METHODOLOGY

The specific interpersonal perception tests we used consist of a series of identical questionnaires that contain statements or adjectives commonly found in personality inventories. For example, we have used statements such as, "I usually avoid close intimacies with others," "I like good food," or paired adjectives such as "intelligent-unintelligent," "calm-excited." Each item is to be marked on a 6-point scale. The subject (S) completes these identical questionnaires under different sets. He is generally asked to describe (a) himself, (b) his ideal-self—how he would ideally like to be, (c) his leader, (d) some person whom he remembers as his best co-worker, and (e) someone whom he remembers as having been his poorest co-worker.

S's questionnaires can then be scored for Assumed Similarity (AS). For example, a person who describes his leader as very similar to himself has high Assumed Similarity to the leader ( $AS_L$ ) while one who describes the leader as quite different from himself has low  $AS_L$ . A person who describes his leader very much as he describes his ideal-self is presumed to see his leader as a more ideal person than one who describes his leader as very unlike his ideal self. Another score with which we shall deal is Assumed Similarity between Opposites ( $AS_o$ ). This score is obtained when we compare a S's descriptions of his most and of his least preferred co-workers.

Assumed Similarity scores are very reliable measures with split half coefficients of .85 to .95 for twenty to thirty item tests. Studies have led to the interpretation of AS scores as measuring on a dimension of acceptance or closeness vs. psychological distance. Thus, it has been demonstrated that a person who is liked or preferred as a work companion is seen as more similar than one who is disliked or not preferred (7, 9). W. Hausman has also shown that Ss feel closer to a parent who is similar than to one who is described as being relatively dissimilar to S.<sup>2</sup> Bieri (1) has recently reported that perceived similarity increases as a S becomes better acquainted with the person who is to be described. In light of these and other findings we interpret high  $AS_L$  as indicating a feeling of closeness or familiarity to the

<sup>2</sup> W. Hausman, unpublished research, West Point Military Academy.

leader, while low  $AS_L$  indicates a feeling that the leader is different from oneself, and hence, a feeling of distance.

The score  $ASo$  (the perceived similarity between the best and poorest co-workers  $S$  has ever had), is similarly interpreted as a generalized index of psychological distance between  $S$  and his co-workers. It indicates the leader's readiness to reject any person who is a poor co-worker, irrespective of other (e.g., personal) considerations (8).

Let us now briefly review the evidence which our studies provide for and against the hypotheses suggested by the familiarity-breeds-contempt position.

a. *The subordinates' feeling of familiarity toward their leader is detrimental to group effectiveness.*

This hypothesis implies that men who feel close to their leader—who perceive him as being very much like themselves—will follow him less readily than men who do not feel close to their leader. The implication is that the good leader must be seen by his followers as a better man or that he must be thought to have special attributes which differentiate him from his followers. Else, why should a person permit his leader to make important decisions for him—even those which affect life and death—if he is no better than any of his followers? Hence, men who perceive their leader as similar to themselves will be a poorer team than men who perceive their leader as different (have low  $AS_L$ ).

As can be seen from Table 1, this hypothesis is not supported: the similarity which members of various teams perceive between themselves and their leader is uncorrelated with team effectiveness.

b. *The men's knowledge of their leader is inversely related to group effectiveness.*

The argument has been advanced that a leader should never show his weak points. If this be true then a leader who is known by his men, and whose weak points are presumably also known, will be followed less readily

TABLE 1

*Relations between Assumed Similarity to the Leader and Team Effectiveness Criteria\**

Sample	Criterion	N	Correlation
B-29 bomber crews (10)	Radar bomb scores	55	.00
Tank crews (10)	Time to travel to and hit 25 targets	25	-.01
Sales cooperatives (12)	Net income	26	.12

\* Space does not permit an extensive summary of each study that will be mentioned in this paper. The interested reader may find details concerning criterion reliability and methodology in the references listed.

than one who is not too well known. (On the other hand, it might also be pointed out that it may be advantageous if the men knew their leader well, especially if the leader had a sterling character.)

Unfortunately, neither proposition can be adequately tested with techniques now at our disposal. The studies on diagnostic ability reported by Kelly and Fiske (14), as well as the methodological studies by Cronbach (4) and Crow (5), clearly indicate that "understanding others"—as measured by the ability to predict their self-concept or their personality attributes—cannot be reliably measured. These investigators come to the reluctant conclusion that a man's ability to predict attributes which others see in themselves is not much better than chance. For the sake of the record, we might mention that ability to predict the responses of the informal leader of a basketball team does not correlate with the criterion of per cent of games won ( $Rho = -.16$ ,  $N = 14$ ). The crew's ability to predict the tank commander's responses correlates  $-.35$  with crew effectiveness ( $N = 22$ ). Whether or not different types of predictions would lead to better results is a matter for further research.

It has also been said that men who know fewer of the leader's weak points will idealize him more, and that men who idealize their leader will work better than those who do not. Hence, we would predict that *idealization of the leader is positively correlated with group effectiveness*. This hypothesis was tested in our study of basketball teams but is unsupported ( $Rho = .17$ ,  $N = 14$ ). In addition, we must point out that idealization of the leader, as measured by comparing S's ideal-self and his description of the leader, is correlated with self-leader correspondence ( $AS_L$ ) almost to the extent of the reliability of the scores (3). Hence, we should obtain essentially the same results as those presented on Table 1.

The hypothesis might also be tested by means of sociometric preference ratings. We generally assume that a popular, highly chosen leader is idealized by his men, as evidenced by the presence of a large halo effect in the men's ratings of the leader. These ratings generally show that a person who is seen as a good leader is also seen as a good problem-solver, as a good therapist, as strong, intelligent, and so on, even though these attributes are uncorrelated and may not, in fact, apply to the leader. This interpretation is also supported by investigations of French (11) and others. The hypothesis has, therefore, been advanced that *highly popular leaders have more effective groups than unpopular leaders*. However, our studies have shown no systematic trends of this nature when we use *objective* group effectiveness criteria. As Table 2 indicates, leaders of effective groups are no more highly chosen than are leaders of less effective groups. This experience is consonant with evidence reported by others who have used objective criteria i.e., those not based on subjective ratings (15). While some positive relations

TABLE 2

*Correlations between Sociometric Preference Scores Received by the Leader and Group Effectiveness Criteria*

Sample	Criterion	N	Correlation
Tank crews (10)	Time to travel to and hit 25 targets	25	.27
Open-hearth foreman groups (2)	Tonnage of steel per unit time	14	-.49
Sales cooperatives (12)	Net income	32	.12
	Operating efficiency		.06

TABLE 3

*Correlation between the Accepted Leader's ASO Score and the Team's Effectiveness Criteria\**

Sample	Criterion	N	Relation†	Correlation measure
Basketball teams (9)	Per cent games won	14	-.69	Rho
Basketball teams	Per cent games won	12	-.58	*p.b.
Surveying parties (9)	Accuracy in mapping	22	-.51	r
Open-hearth Shop (2)	Heat time (tonnage per unit of time)	15	-.52	Rho
Sales cooperatives (12)	Net income			
where general manager is accepted by:				
Board's leader		23	-.39	Rho
Both board's leader and staff		13	-.70	Rho

\* A special case of this principle is presented by rigidly structured military units in which personnel turnover is extremely high. Here we find that the accepted leader's psychological distance manifests itself either by assuming low similarity or by sociometrically not choosing his key subordinates. This complex interaction is discussed in the research report on this study (10).

† ASO is defined as the perceived *similarity* between most and least preferred co-workers. The negative correlations indicate, therefore, that leaders of effective groups see greater differences between most and least preferred co-workers than do leaders of the less effective groups.

have been reported (e.g., 16), these must be viewed in light of equally impressive negative findings.

We thus turn to our alternative proposition which holds that the customs under discussion are designed to prevent the leader from becoming too emotionally involved or familiar with his men. Hence, we derive the hypothesis that *leaders who perceive themselves to be more distant from their sub-*



*ordinates have more effective work units than those who perceive themselves to be closer to their men.*

As we indicated above, the index ASo, the Assumed Similarity between the most and the least preferred co-workers, has been interpreted as a generalized measure of perceived psychological distance to one's work companions.

Table 3 presents the correlations between the leader's ASo score and group effectiveness criteria.

Informal leaders, with low ASo scores, of basketball teams and surveying parties had more effective work groups than those with high ASo scores (8). Likewise, studies on open-hearth foremen (2) and managers of sales cooperatives (12) showed that the psychologically more distant leaders had more productive and efficient units than leaders who tended to perceive their co-workers as emotionally closer to them.

This finding is also in accord with that reported by Katz, Maccoby, Gurin, and Floor (13) in a study of railway section gangs. These investigators report that "low foremen" do not clearly perceive their leadership role. "High foremen are typically more aware of their position as leader and supervisor. . . ." We take this statement to mean that these high foremen do not try to be "one of the boys" even though they may be concerned with the welfare and personal well-being of their men.

In addition, Shaw's laboratory study on communication networks (18) indicates that groups with autocratic (and presumably more distant) leaders performed more effectively than groups with nonautocratic (democratic) leaders. This study also suggests that the more distant, autocratic leadership situation protects the leader by restricting the incoming message units with which he might have to deal.

#### SUMMARY

In summary, our studies throw some light on the customs of military and industrial organizations which require, or strongly encourage, the maintenance of social distance between the leader and his subordinates. Our studies support the hypothesis that these customs increase group effectiveness not so much because the men will not follow a leader with whom they are too familiar, but rather because a leader, who is too close to his men, might find it more difficult to reach decisions which are uninfluenced by his feelings about the men.

*Manuscript received: August 24, 1966*

*Revised manuscript received: December 7, 1966*

*Fred E. Fiedler*

*Department of Psychology*

*University of Illinois*

*Champaign, Illinois*



## REFERENCES

1. Bieri, J., "Changes in Interpersonal Perception Following Social Interaction," *Journal of Abnormal and Social Psychology*, 1953, 48, 61-66.
2. Cleven, W. A., and F. E. Fiedler, "Interpersonal Perceptions of Open Hearth Foremen and Steel Production," *Journal of Applied Psychology*, 1956, 40, 312-314.
3. Cronbach, L. J., W. Hartmann, and M. E. Ehart, "Investigation of the Character and Properties of Assumed Similarity Measures," (mimeographed paper) Technical Report No. 7, Contract N6-ori-07135, Urbana, Illinois: Bureau of Research and Service, University of Illinois, 1953.
4. Cronbach, L. J., "Processes Affecting Scores on 'Understanding of Others' and 'Assumed Similarity'," *Psychological Bulletin*, 1955, 52, 177-193.
5. Crow, W. J., "A Methodological Study of Social Perceptiveness," Ph.D. dissertation, University of Colorado, 1954.
6. Dorn, W. L., *Competition for Empire*, New York: Harper, 1940.
7. Fiedler, F. E., W. G. Warrington, and F. J. Blaisdell, "Unconscious Attitudes as Correlates of Sociometric Choice in a Social Group," *Journal of Abnormal and Social Psychology*, 1952, 47, 790-796.
8. Fiedler, F. E., "The Psychological Distance Dimension in Interpersonal Relations," *Journal of Personality*, 1953, 22, 142-150.
9. Fiedler, F. E., "Assumed Similarity Measures as Predictors of Team Effectiveness," *Journal of Abnormal and Social Psychology*, 1954, 49, 381-388.
10. Fiedler, F. E., "The Influence of Leader-Keyman Relations on Combat Crew Effectiveness," *Journal of Abnormal and Social Psychology*, 1955, 51, 227-235.
11. French, R. L., "Sociometric Status and Individual Adjustment among Naval Recruits," *Journal of Abnormal and Social Psychology*, 1951, 46, 64-73.
12. Godfrey, E. P., C. F. Wrigley, B. F. Mannheim, D. M. Hall, and F. E. Fiedler, "The Effect of Interpersonal Relations on the Success of Consumer Cooperatives," *Abstracts of Papers Delivered at the 50th Annual Meeting of the American Sociological Society*, (mimeographed paper) American Sociological Society, 1955.
13. Katz, D., N. Maccoby, G. Gurin, and L. G. Floor, *Productivity, Supervision and Morale among Railroad Workers*, Ann Arbor, Michigan: University of Michigan Press, 1951.
14. Kelly, E. L., and D. W. Fiske, *The Prediction of Performance in Clinical Psychology*, Ann Arbor, Michigan: University of Michigan Press, 1951.
15. Palmer, F. H., and T. I. Myers, "Sociometric Choice and Group Productivity Among Radar Crews," (abstract) *American Psychologist*, 1955, 10, 441.
16. Roby, T. B., "The Relationship between Sociometric Measures and Performance in Medium Bomber Crews," *Human Resources Research Center Research Bulletin 53-41*, Lackland Air Force Base, 1953.
17. Rosinski, H., *The German Army*, New York: Harcourt Brace, 1940.
18. Shaw, M., "A Comparison of Two Types of Leadership in Various Communication Nets," *Journal of Abnormal and Social Psychology*, 1955, 50, 127-134.

## The Behavior of Emergent and Designated Leaders in Situational Tests<sup>1</sup>

WALTER R. BORG, *Maxwell AFB, Alabama*

The general aim of this research was to determine (a) how individual and group performance in situational test problems changes when a leader is designated after an opportunity has been presented for a natural leader to emerge in previous problems and (b) to compare performance of groups in which a leader emerged with groups in which no leader emerged and groups in which two leaders emerged. The specific hypotheses to be tested by this research are as follows:

- a. The effectiveness of the emergent leader is reduced when another group member is designated leader.
- b. In a six-man group in which a leader has emerged, an increase in negative behavior takes place when another group member is designated leader.
- c. The designated leader (when other than the emergent leader) shows more leadership behavior than in problems where he is not designated leader.
- d. In groups where two leaders have emerged, the emergent leader displays more leadership behavior when he is also the designated leader.
- e. Followers in groups having an emergent leader make higher behavior scores than members of groups in which no leader has emerged.
- f. The amount of desirable leadership behavior of the team as a whole is positively related to the behavior score of the designated leader.

### METHOD

Students of the USAF Officer Candidate School are selected on the basis of intelligence, education, time in service, and enlisted grade. Approximately 11 per cent of those applying are accepted for the school. All male members of two complete classes at the USAF Officer Candidate School were subjects in this research. Upon reporting to the Officer Candidate School, students were assigned to six-man teams. Class 1955-B was divided into 22 six-man teams. Four of the teams could not be scheduled for all problems, leaving a total of 18 teams that completed all problems. Class 1955-C

<sup>1</sup> This report is based on work done under ARDC Project No. 7719, Task 17009, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

entered training three months later. This class was divided into 23 six-man teams all of which completed all problems. Teams were arranged so that no member of a team had been acquainted with any other member of the team prior to reporting to the school.

The test consisted of twelve situational problems. These problems all require cooperation of team members in developing and carrying out a solution and provide a situation in which leadership behavior can be displayed. In a typical problem, the team is placed in a prison compound and must escape across a moat and over a solid board fence. A ladder is provided, but it is not long enough to reach from the edge of the moat to the top of the fence. Several short lengths of rope are also provided. Certain areas, such as the edge of the moat, are painted red and must not be touched. The problem is solved by using the ropes to hold the ladder at an angle over the moat. One team member climbs the ladder and jumps to the fence. Once over the fence he discovers additional props which can be used to help the remainder of the team to escape. A time limit of twelve minutes was imposed on each problem which is usually insufficient to complete the problem. Solution of the problem was not considered essential as the rating scale used was directed toward behavior of team members rather than solution of the problem. Every precaution was taken to maintain isolation of the teams during the testing period. All subjects were informed at the initial briefing that they were not to speak to anyone in their class except members of their own team. A strict honor system in effect at the school, plus close supervision by designated members of the upper class, kept communication between members of different teams to a minimum. The problems were administered by Air Force officers attached to the Officer Candidate School. Two officers were assigned to each team to administer the first six problems. At the end of six problems, the officers were moved and the last six problems were administered by two other officers. Six upperclassmen were assigned to each team to act as raters. All raters were given training in the procedures prior to the research, but upperclassmen rating Class 55-C were somewhat better qualified since they had gone through all the problems as lowerclassmen. An upper-class rater was assigned to each individual in the team and rated only the individual assigned to him for a given problem. At the end of each problem, raters were rotated. This system resulted in each team member being rated twice by each rater in the course of completing the twelve problems. In order to compensate for differences in difficulty level of the twelve problems, team A started with problem 1, team B with problem 2, and so on. The rating device employed was a behavior check list consisting of the following eleven categories:

1. *Initial planning:* Figuring out a solution with major steps in sequence

(student must communicate this plan to some team member to receive credit).

2. *Defining the problem*: Pointing out aspects of the problem or solution that have been overlooked.

3. *Making suggestions*: Proposing an approach or procedure.

4. *Communicating*: Spreading information to other team members. Rephrasing and explaining.

5. *Coordinating*: Attempting to compromise between conflicting solutions or suggestions.

6. *Initiating discussion*: The first person to speak after a lull in talk.

7. *Initiating action*: The first person to move into a specific task (before being told to do so).

8. *Negative reaction*: Displays anger, disgust, or distaste toward other members of the group.

9. *Distracts*: Takes the group's attention away from the problem.

10. *Criticizes*: Criticizes ideas, procedures of the leader or other team members. Argues against solution adopted by the group.

11. *Seeking attention*: Uses various devices to direct attention toward himself.

The rater was instructed to place a check for the man he was assigned to observe each time the man displayed one of the behaviors covered in the check list. For example, a man who made three suggestions during the problem would be given three checks. Raters were cautioned to make checks only when a specific observable behavior incident occurred. They were cautioned not to make general ratings or trait ratings. The rating areas were developed by the author and based upon extensive observation of teams working on the twelve problems prior to the start of this research. These areas overlap considerably with the rating areas employed by other psychologists and sociologists to measure small group leadership. The subject's leadership behavior score is arrived at by taking the sum of all checks given in categories 1 through 7 minus checks given in categories 8 through 11. Very few negative checks (8 through 11) were given. Separate behavior scores were computed for the first six and second six problems in which the team participated. In the first six problems no leader was designated. In the second six problems the introduction to the problem (which sets the scene) specified a particular team member as being the senior officer or leader in the situation. Each team member was designated leader in one problem. In the initial briefings before the problems were started, subjects were told that the purpose of the problems was to evaluate the ability of the team to work together in solving the problem. Although motivation is generally high upon entry into Officer Candidate School, additional motivation was provided by offering a special pass to members of the team that did the best job of solving the twelve problems.

## RESULTS

In analyzing the data it was found that there were differences in means, standard deviations, and reliabilities between ratings of Class 55-B and Class 55-C. It was decided, therefore, to treat the data for the two classes separately whenever possible. In the analysis of teams exhibiting special characteristics (such as two competing leaders) it was necessary to draw from both classes in order to have a sufficient number of cases. Limited comparisons between these special groups are possible, however, in spite of class differences as the special groups contained about the same proportion of teams from each class.

Behavior scores of all team members of the 41 teams were studied. In 17 of these teams, scores on the first six problems indicated that a leader had emerged. The criterion for an emergent leader was a behavior score on the first six problems in the upper 20 per cent and one-half standard deviation above the next highest score in the team. In a total of 14 teams, no clear-cut leader emerged in the first six problems. In the remaining 10 teams two leaders emerged, both of whom earned markedly higher behavior scores than other team members but were not significantly different from each other.

The first portion of the analysis is concerned with behavior of the 17 teams in which one leader clearly emerged during the first six problems. One aim of this phase of the study was to determine whether changes in the behavior score of the emergent leader occurred during the second six problems in which some other group member was designated as leader. In order to determine the degree of change, the behavior score of the emergent leader on the first six problems, in which no leader was designated, was compared to his behavior score on five of the last six problems, omitting the problem in which the emergent leader was also designated leader, and multiplying his score on the five remaining problems by 1.20. Comparison of these two scores showed that the emergent leader's score was higher in the first six problems in 15 out of 17 cases. The mean behavior score of the emergent leaders on the first six problems was 42.1 ( $\sigma = 12.78$ ) versus a corrected mean behavior score of 30.7 ( $\sigma = 11.39$ ) on the latter problems. The correlation between the behavior score on the first six and the second six problems was .72 for Class 55-B and .73 for Class 55-C. The average score of the emergent leaders on the latter problems would be expected to regress to 34.6. The score one would expect on the second six problems, using the regression equation, was significantly higher at the 1 per cent level than the actual score made by the emergent leaders ( $CR = 5.65$ ).<sup>2</sup> The emergent leader was also designated leader on one of the last

<sup>2</sup> All critical ratios reported in this study are based on the result of dividing the difference between means by the standard error of the difference between means.

six problems. The average corrected score of the emergent leaders on this problem was 34.6, which is the score predicted by the regression equation. It is interesting to note that the emergent leaders' mean score on the five problems where another leader is designated is still significantly higher than the mean behavior score of other members of these teams which was 22.83.

We have seen above that the performance of the emergent leader in the second six problems is significantly lower than in the first six problems where no leader is designated. It might be expected that the reduction in leadership behavior among the emergent leaders would be compensated for by an increase in leadership behavior of designated leaders in the last six problems. Several studies have found that appointing or designating an individual as leader usually results in an increase in the leadership behavior rating of the designated individual. In a study concerned with the effect of control on leadership, Bass found that an increase in control for a given individual resulted in an increase in attempted leadership acts (1). Carter, in a similar study, indicated that the power attending the appointment of an individual as leader resulted in the individual receiving significantly higher leadership ratings than the average for other group members even though these appointed leaders were introduced into an established group (3). In order to determine whether an increase in leadership behavior was evidenced in the present research, the total behavior score of each team member, other than the emergent leader, on the first six problems was compared with his score on the problem where he was designated leader. The results of this comparison showed no significant difference in behavior on the part of individuals when they were designated as leaders. The mean score for the 83 individuals involved on the first six problems was 22.83 as compared with a corrected mean of 22.55 for the problem in which the individual was designated leader. Thirty-eight individuals received higher scores when they were designated leaders, 41 received lower scores when designated leaders, and 4 remained the same.

It was hypothesized that in the 17 teams in which a leader had already emerged the designation of another individual as leader would result in an increase in negative behavior on the part of the team members. This hypothesis was not borne out by results based on ratings of negative behavior. In order to make an estimate of negative behavior, the total check marks on the four negative items on the behavior rating sheet were totaled. Comparison was made between the number of negative checks on the first six problems and the number of negative checks on the last five problems, omitting the problem in which the emergent leader was also designated leader. The score on the last five problems was then multiplied by 1.2 in order to make it roughly comparable to the score on the first six problems. Total negative checks on the first six problems was 73 as compared to the corrected total of 70.8 on the last five problems.



In 23 of the 41 teams, one individual stood out from the other team members as being markedly inferior in terms of the leadership behavior score. These individuals were in the bottom 20 per cent in terms of total behavior score and in most cases were much lower than any other team member. In order to determine the effect of poor leadership upon total team performance, the total behavior score of each team on the problem where the low man was designated leader was computed. The mean team score when the lowest team member was designated leader was 16.96. The mean team score of these same teams on the problems where the low man was not designated leader was 18.58. This difference was not significant. Total negative checks when these individuals were designated leader was not significantly increased.

In 14 teams no leader clearly emerged in the first six problems. The average score of individuals on these teams for the second six problems was significantly lower than scores made by individuals on teams where one or two leaders emerged ( $CR = 4.82$  vs. one-leader teams;  $CR = 2.88$  vs. two-leader teams). These differences are partially spurious because a high score on the first six problems is one of the criteria for selecting emergent leaders, and scores on the first six problems correlate .72 with scores on the second six problems. To remove the spurious element the scores of the emergent leaders were subtracted from the 17 teams in which one leader had emerged. The average score of the remaining team members was significantly higher ( $CR = 3.75$ ) than the score made by members of teams in which no leader clearly emerged. Thus, the followers on well-led teams showed significantly more desirable behavior in the areas rated than the average for leaderless teams. The above finding is in agreement with other research involving small groups. A study by Haythorn, for example involving two-man groups indicates that a person of high performance facilitates the performance of the person with whom he works (4). The "great men" described by Borgatta, Bales, and Couch were selected by methods similar to those used for selecting emergent leaders in the present research (2). It was found that the product index of agreement and suggestion, which is indicative of favorable group behavior, was significantly higher for groups in which "great men" participated.

In 10 of the 41 teams studied, two leaders emerged during the first six problems, both of whom earned markedly higher scores than other team members but were not significantly different from each other. A study of the behavior of these 20 individuals showed an interesting relationship between their scores in the leaderless situation when designated leader and when the competing emergent leader was designated leader. The mean score of the 20 emergent leaders on problems where neither was designated leader was 27.22. The mean score of the 20 emergent leaders on the problem where



each was designated leader was 34.4, while the mean score of the emergent leader when the competing emergent leader on the team was designated leader was only 26.1. This difference approached significance ( $CR = 1.74$ ).

A comparison of the performance of these ten teams on the first six problems versus the second six problems shows a drop in the mean behavior score in the second six problems in which a leader was designated. The mean individual score on the first six problems was 24.31 as compared with a mean of 21.71 for the second six problems. The difference between these two means is not significant, however ( $CR = 1.65$ ).

In order to determine whether the amount of desirable behavior of the team as a whole is related to the leadership behavior score of the designated leader, each individual's total leadership behavior score was correlated with the total score of the other team members in the problem where the individual was designated leader. For Class 55-B this correlation was only .14, but in Class 55-C the correlation was .54. This difference is accounted for in part by the better reliability of ratings for Class 55-C (83 as compared with .66 for Class 55-B). It may also be assumed that the ratings for Class 55-C are also somewhat more valid. It will be remembered that Class 55-C was rated by members of Class 55-B who had been through all the problems and had a better insight into the entire evaluation than the students in Class 55-A who conducted the first rating.

#### SUMMARY

The objective of this research was to determine how individual and group performance changes when a leader is designated after an opportunity has been presented for a natural leader to emerge in previous problems and to compare performance in groups in which a leader emerged with groups in which no leader emerged and groups in which two leaders emerged. All members of 41 six-man teams taken from OCS Classes 55-B and 55-C were rated on the basis of leadership performance in 12 situational problems. On the basis of performance on the first six problems the teams were divided into three groups: 17 teams in which a leader had clearly emerged; 14 teams in which no leader had clearly emerged; and 10 teams in which two competing leaders had emerged. The performance of individual team members in these three groups was analyzed and a comparison between the three groups was made. The following conclusions are based on this analysis:

1. The average effectiveness of the emergent leader is significantly reduced when a team member other than the emerged leader is designated as leader ( $CR = 5.65$ ).
2. The emergent leader's score when another leader is designated, however, is still significantly higher than the mean behavior score of other team members ( $CR = 2.21$ ).

3. The performance of designated leaders (other than the emergent leader) is not significantly different than the individual's performance when he is not designated leader. The lower efficiency of the emergent leader, therefore, is not compensated for by an increase in efficiency of the designated leader.

4. When the emergent leader is designated leader, his leadership behavior score is not significantly different from the score predicted by the regression equation.

5. There is no significant increase in negative behavior on the part of team members when an individual other than the emergent leader is designated leader. The small use by raters of the negative items on the rating scale, however, suggests that these items were not a satisfactory indicator of negative behavior.

6. When individuals who are markedly inferior in terms of leadership behavior score are designated leaders the mean team score is not significantly lower than when other team members are designated leaders.

7. Negative team behavior is not significantly increased when an individual, who is markedly inferior in terms of behavior score, is designated leader.

8. The over-all performance of teams in which no leader emerged is significantly inferior to the performance of teams in which one leader has emerged ( $CR = 4.82$ ) and teams in which two leaders have emerged ( $CR = 2.88$ ).

9. Followers on teams in which a leader has emerged make significantly higher mean leadership behavior scores than the average for teams where no leader has emerged ( $CR = 3.75$ ). This suggests that the presence of a leader increases leadership performance of followers rather than suppressing such performance as has sometimes been assumed.

10. Teams in which two leaders have emerged make significantly lower average leadership scores than teams with one leader ( $CR = 1.97$ ).

11. In comparing leadership scores of the 20 competing leaders, it was found that the mean score on the problem in which the individual was designated leader was not significantly higher than the mean score on the problem where competing leader was designated leader ( $CR = 1.74$ ). The magnitude of the differences in mean scores, however, suggests that more extensive research might show a significant difference.

12. Study of the performance of teams having two emergent leaders revealed that mean scores on the first six problems were higher than scores on the second six problems in which a leader was designated. This difference, however, was not significant ( $CR = 1.65$ ).

13. In Class 1955-C there was a significant relationship between the leadership score of the designated leader and the performance of the other

members of the team while under his leadership ( $r = .54$ ). This relationship was much lower in Class 55-B ( $r = .14$ ). This difference is partially accounted for by the lower reliability of ratings for this class. As raters for Class 55-B had not been through the test problems themselves, they almost certainly had less insight into the situations, and this fact may also have lowered the validity of their ratings.

#### IMPLICATIONS

It appears to the author that this study has two implications with respect to leadership behavior in small groups. The first of these is that the designation of the wrong individual as leader in a small group situation does little to improve this individual's effectiveness but tends to suppress the emergence of a more capable leader. It would be desirable to test the results of this study in situational problems where the designation of a leader could be accomplished at a more realistic level. A study in which officer teams were used with the highest ranking officer designated leader might accomplish this.

A more important implication, however, which the author feels can be accepted with somewhat less reserve, hinges on the degree of relationship between the ability of the leader designated and total performance of the team found in this study. The designation of a competent leader appears to be a major determiner of team effectiveness and seems to stimulate effective leadership behavior among other members of the team rather than suppress it.

There is considerable evidence that future military operations will revolve to an ever-increasing extent around small teams of specially trained individuals. The increase in highly specialized weapons systems, and the necessity for developing moderately independent and highly decentralized defensive and retaliatory forces in the atomic age, are major factors in this trend. Organizing optimum small groups and selecting the best small group leader may well be the most important personnel problem in the armed forces during the next 20 years.

*Manuscript received: October 29, 1956*

*Revised manuscript received: January 18, 1957*

Walter R. Borg

Mazwell Field Unit Number 1

Personnel Laboratory, AFPTRC

Mazwell Air Force Base, Alabama

#### REFERENCES

1. Bass, B. M., E. L. Gaier, and A. W. Flint, *Attempted Leadership as a Function of Motivation Interacting with Amount of Control*, Technical Report 7, Washington: Office of Naval Research, 1956.

2. Borgatta, E. F., R. F. Bales, and A. S. Couch, "Some Findings Relevant to the Great Man Theory of Leadership," *American Sociological Review*, 1954, 19, 755-759.
3. Carter, L. F., W. Haythorn, B. Shriver, and J. Lanzetta, "The Behavior of Leaders and Other Group Members," *Journal of Abnormal and Social Psychology*, 1951, 46, 589-595.
4. Haythorn, W., "The Influence of Individual Members on the Characteristics of Small Groups," *Journal of Abnormal and Social Psychology*, 1953, 48, 276-284.

## Anomie, Social Isolation, and the Class Structure<sup>1</sup>

WENDELL BELL, *Northwestern University*

Several recent studies contain data which suggest that anomie, as measured by the Srole scale, is an important determinant of prejudice (7, 12, 13). It is less clear, however, as to what factor or factors may be important in producing anomie in an individual. Srole has found economic status to be negatively related to anomie, but other researchers interpret their data to the contrary.<sup>2</sup> There remain, however, persuasive theoretical considerations which would lead one to contend with Merton that the strain toward anomie operates unevenly throughout the various segments of the social structure, with persons in low economic positions being characterized by personal demoralization or disorganization (6). Thus, one purpose of this paper is to bring a new set of data to bear on the hypothesis that economic status is an important determinant of anomie.

A second purpose of the paper is to locate other factors which may be related to anomie. The most important of these considered here is social isolation, the hypothesis being that those persons who are relatively isolated are more likely to be anomic than persons who are less isolated. Participation in informal groups such as the neighborhood group, the kin group, the work group, and the friendship group as well as participation in formal groups can be expected to be inversely related to anomie in that the individual achieves through such participation not only close, intimate, and personally satisfying social relations, but also a sense of being a meaningful and effective part of the society.

The third purpose of this paper is to relate both unit and personal measures of economic status to anomie. The economic characteristics of the neighborhood in which a person lives (a unit variable) is expected to affect his responses to the anomie scale even when his individual economic characteristics (a personal variable) are controlled.

<sup>1</sup> A revised version of a paper read at the 33rd Annual Institute of the Society for Social Research of the University of Chicago, June, 1956. The writer wishes to acknowledge the financial assistance given him by the Carnegie Corporation of New York and the Stanford University Committee for Research in the Social Sciences for the execution of the study of which this paper is part. Marion D. Boat, Maryanne T. Force, and Richard K. Pumilia made important contributions to this paper.

<sup>2</sup> See the exchange between Srole (11) and Rokeach (8) as well as their original papers.

## DESIGN OF RESEARCH

The data given below were collected in the spring of 1953 in connection with a study of social participation conducted in four San Francisco census tracts. The tracts (or neighborhoods) were selected according to their scores on the Shevky urban typology (9, 10) so that they would contain few nonwhites or foreign-born whites from certain countries, and so that they would vary widely with respect to economic and family characteristics. The economic status of a tract population is measured by a composite score based on occupation and education; the family status of a tract population is measured by a composite score based on fertility, per cent of women not in the labor force, and the per cent of dwellings which are single-family detached. The four tracts studied are: the Mission district, which is a low rent rooming-house area having a low score on the index of economic status and a low score on the index of family status; Outer Mission, which is low on economic but high on family status, being a low rent single-family detached home area; Pacific Heights, a high-rent apartment-house area, which is high on economic and low on family status; and St. Francis Wood, which is high on economic and family status, being characterized by large and expensive single-family detached homes.

In each of these four neighborhoods probability area samples were selected. The respondents were men aged twenty-one or over, and the total response rate was in excess of 85 per cent. A total of 701 interviews were obtained.<sup>3</sup>

## INDEXES EMPLOYED

*Srole's Anomie Scale*

As indicated above, the measure of anomie used in this study is the anomie scale constructed by Leo Srole (12, 13). According to Srole, this scale was constructed to measure the phenomena variously referred to as social dysfunction or disorganization, group alienation, demoralization, or what Durkheim called anomie. The scale consists of five questions as originally constructed, and Srole describes in detail what each is intended to measure. The first was constructed to measure the individual's sense that community leaders are detached from and indifferent to his needs. The second is designed to measure the individual's perception of the social order as essentially fickle, unpredictable, and orderless, giving rise to the feeling that he could do little to direct his life with any degree of time perspective or planning ahead. The third is purported to measure the individual's view that he and other people like him are retrogressing from the goals they have already reached. The fourth is an attempt to measure the deflation or loss of meaning of internalized group norms, values, and goals, resulting in

<sup>3</sup> For a detailed statement of the research design employed, see Bell (1).

extreme form in the individual's sense of the meaninglessness and futility of life itself. And the fifth is predicted to measure the individual's sense that the framework of immediate personal relationships is no longer predictive or supportive. In this paper the terms "anomie," "disorganization," "demoralization," or "personal morale" stated conversely, will refer to the attitude measured by Srole's items.<sup>4</sup>

The questions comprising the anomie scale were asked of the respondents in each of the census tracts. Although only five items are in the scale, the Cornell scaling technique was applied. The coefficient of reproducibility for this sample is .90 and the coefficient of scalability is .65, providing some validity in each case to the assumption that the items constitute a unidimensional scale. Scaling the items resulted in two trichotomies and three dichotomies, the scale scores ranging from a possible low of zero to a possible high of 10.<sup>5</sup>

#### *Measures of Social Participation*

The measures of social participation used include the frequency of participation in certain informal groups, formal group membership, and formal group attendance. The respondents were queried about the frequency of their participation with their neighbors, their relatives other than those living with them, their co-workers outside of work, and friends other than neighbors, relatives, and co-workers. These data have been analyzed elsewhere and it was concluded that for the most part these informal social relations, when they occur, are likely to represent close, primary, and intimate associations (4). For the purposes of this paper the frequency of participation with each of these groups was summed to yield a single frequency indicating the total informal participation of the respondents in all these groups combined.

In addition to total informal group participation, memberships and attendance of meetings of formal associations were obtained. The definition of formal associations follows that of other writers studying voluntary associations, formal groups, etc. Business concerns and governmental agencies are excluded from the types of formal groups considered here. These data have also been discussed elsewhere (2, 3).

#### *Indexes of Economic Status*

To measure the economic status of the census tract populations, occupation and education were used. In each tract the percentage of employed

<sup>4</sup> The reader is referred to Srole's published paper (13) which is a definitive draft of his earlier paper (12) and contains his most recent thinking on his anomie scale. Srole there introduces the term "anomia" to refer to his scale.

<sup>5</sup> It should be noted that contrary to the results of this study, the anomie items failed to constitute a unidimensional scale in a recent study of a small, homogeneous, upper-middle-class, suburban area (14).



persons who are craftsmen, operatives, and laborers, and the percentage of persons over age twenty-five who have only a grade-school education or less, were determined. After standard scores were computed these were summed to give an economic status score. As indicated above, by this index Pacific Heights and St. Francis Wood have populations characterized by high economic status and Mission and Outer Mission contain populations having low economic status. This score will be referred to as *neighborhood economic status*.

In addition, each respondent was given an economic status score based upon his own occupation, education, and income. This measure will be referred to as *individual economic status*.

#### RESULTS

Table 1 shows the mean anomie scale scores for each of the four study neighborhoods. It was hypothesized that persons living in Mission and Pacific Heights would have higher anomie scores on the average than persons living in Outer Mission and St. Francis Wood respectively on the basis of the differences in family life among these neighborhoods. Mission and Pacific Heights contain populations having many unrelated individuals, many women in the labor force, and relatively few children, while Outer Mission and St. Francis Wood contain populations having relatively many persons living in family units, few women in the labor force, and relatively many children. However, this hypothesis is not supported by the data. There is no difference between Mission and Outer Mission, and the mean difference in anomie scores between Pacific Heights and St. Francis Wood is not significant.

Comparing Mission with Pacific Heights and Outer Mission with St. Francis Wood, however, supports the hypothesis that persons living in low economic status neighborhoods are more personally demoralized than persons living in high economic status neighborhoods. The mean differences in each case are significant beyond the .001 level. Men living in neighbor-

TABLE 1  
*Mean Anomie Scale Scores by Neighborhood*

Low Econ. Low Family (Mission)	High Econ. Low Family (Pacific Hts.)	Low Econ. High Family (Outer Mission)	High Econ. High Family (St. Francis Wood)
4.5	3.2	4.5	2.8

Significant differences:

Mission vs. Pacific Hts.\*

Outer Mission vs. St. Francis Wood.\*

\*  $p < .001$ .

hoods characterized by low educational and income levels and occupations such as craftsmen, operatives, and laborers have significantly higher anomie scores than men living in neighborhoods characterized by high educational and income levels and white-collar occupations—especially professionals, proprietors, managers, and officials.

In Table 2 the neighborhoods have been grouped according to their economic status, and mean anomie scale scores are given not only by neighborhood economic status but also by frequency of informal group participation. Comparing those men who are relatively isolated from informal group contacts with those who have reasonably frequent informal group participation, it can be seen that the isolates have a higher mean anomie score than the participants in the low economic status neighborhoods ( $p < .01$ ). However, frequency of informal group participation apparently does not affect personal morale in the high economic status neighborhoods, since the mean anomie scores between isolates and participants in these neighborhoods are not significantly different. Thus, the hypothesis that persons who participate seldom or never in informal groups will be more anomic than persons who participate relatively frequently in these groups is borne out only among those men living in low economic status neighborhoods.

One explanation for this might be found in the wide variation in formal group membership and attendance associated with the economic status of the neighborhood. Other analyses have shown that the men living in the high economic status neighborhoods belong to many more formal groups and attend formal group meetings much more often than men living in the low economic status neighborhoods (3). Thus relative isolation from

TABLE 2  
*Mean Anomie Scale Scores by Frequency of Informal Group Participation and Neighborhood Economic Status*

Informal Group Participation	Low Economic Status (Mission & Outer Mission)	High Economic Status (Pacific Hts. & St. Francis Wood)
Frequent participation (a few times a month or more).....	4.3	3.0
Infrequent participation (about once a month or less).....	5.4	3.3

Significant differences:

Frequent vs. infrequent participation at low economic status.\*

Low vs. high economic status at frequent participation.\*\*

Low vs. high economic status at infrequent participation.\*\*

\*  $p < .01$ .

\*\*  $p < .001$ .

the informal groups studied here may have no effect on personal demoralization among the men living in high economic status neighborhoods since most of them have reasonably frequent association with others through their formal group memberships. In another paper (4) it was pointed out that most of the men belonging to formal groups have many (nine or more) of their close personal friends among the members of their formal groups; thus, formal group participation might reduce anomie not only by providing the individual with a sense of belonging, group identification, and meaningful and effective social action, but also by providing the individual with close, intimate, and primary type associations as well.

That a correlation exists between anomie and formal group participation can be seen from Table 3. Nonmembers in general have higher anomie scores than members, although this difference is not quite significant at the .05 level in the high economic status neighborhoods. When comparing active with inactive members, the inactives have a higher mean anomie score than the actives in the high economic status neighborhoods ( $p < .001$ ), but no such difference exists between these groups in the low economic status neighborhoods.

These findings, relating informal and formal participation to anomie, are not so convincing as one would desire, but they offer some evidence in favor of the hypothesis that social isolation may result in personal disorganization.

Note that in both Tables 2 and 3 the differences in mean anomie scores

TABLE 3  
*Mean Anomie Scale Scores by Formal Group Participation and Neighborhood Economic Status*

Formal Group Participation	Low Economic Status (Mission & Outer Mission)	High Economic Status (Pacific Hts. & St. Francis Wood)
Nonmembers	5.0	3.5
Members	4.3	2.9
Frequent participation (every other month or more)	4.3	2.7
Infrequent participation (a few times a year or less)	4.4	3.7

Significant differences:

Members vs. nonmembers at low economic status.\*

Frequent vs. infrequent participation at high economic status.\*\*

Low vs. high economic status among nonmembers.\*\*

Low vs. high economic status among members.\*\*

Low vs. high economic status among frequent attenders.\*\*

\*  $p < .05$ .

\*\*  $p < .001$ .

between low and high economic status neighborhoods still exist even when informal group participation and formal group participation respectively are held constant. That is, even within groups of men reporting similar frequencies of social participation the men living in low economic status neighborhoods have higher anomie scores than men living in high economic status neighborhoods.

Other variables which might explain the differences in anomie scores by neighborhood economic status were introduced into the analysis. Table 4, for example, contains mean anomie scores by neighborhood economic status and the residence background of the respondent. The mean differences in anomie between persons having different residence backgrounds are not significant, but within each residence background category the men living in the low economic status neighborhoods still have significantly higher anomie scores than the men living in the high economic status neighborhoods.

Mean anomie scores are given by age and neighborhood economic status in Table 5. Anomie increases with age, the youngest age group having significantly lower anomie scores than the oldest age group in each set of neighborhoods. Upon investigation it does not appear that differential unemployment or retirement between the age groups accounts for the differences in anomie by age. Nor does it appear that either differential informal or formal participation accounts for it; formal participation, for example, increases rather than decreases with age in the high economic status neighborhoods. This difference might be explained, however, by a

TABLE 4  
*Mean Anomie Scale Scores by Rural-Urban Background and Neighborhood Economic Status*

Rural-Urban Background	Low Economic Status (Mission & Outer Mission)	High Economic Status (Pacific Hts. & St. Francis Wood)
Brought up:		
On a farm.....	4.9	3.5
In a small town.....	4.5	2.9
In a middle-sized city.....	4.5	3.0
In a large city.....	4.2	3.0

Significant differences:

Low vs. high economic status among those with farm background.\*

Low vs. high economic status among those with small town background.\*

Low vs. high economic status among those with middle-sized-city background.\*

Low vs. high economic status among those with large city background.\*\*

\*  $p < .01$ .

\*\*  $p < .001$ .

TABLE 5  
*Mean Anomie Scale Scores by Age and Neighborhood Economic Status*

Age	Low Economic Status (Mission & Outer Mission)	High Economic Status (Pacific Hts. & St. Francis Wood)
21-39	4.0	2.6
40-59	4.5	2.9
60 and over	5.4	3.6

Significant differences:

Youngest vs. oldest age group at low economic status.\*\*

Youngest vs. oldest age group at high economic status.\*

Low vs. high economic status among youngest age group.\*\*

Low vs. high economic status among middle age group.\*\*

Low vs. high economic status among oldest age group.\*\*

\*  $p < .01$ .

\*\*  $p < .001$ .

tendency for older aged persons to evaluate their present situation and future prospects in terms of some idealized conception of their past situation. Four of the five anomie questions explicitly refer to a gloomy outlook for the future or a negative view of the present. Omitting item number 1, these are:

2. *Nowadays* a person has to live pretty much for today and let tomorrow take care of itself.

3. In spite of what some people say the lot of the average man is *getting worse*.

4. It's hardly fair to bring children into the world *with the way things look for the future*.

5. *These days* a person doesn't really know who he can count on. [Italics mine.]<sup>6</sup>

Of course, whether or not this tendency to evaluate the past more highly than the present is based upon an idealized conception or a realistic assessment of the past and present remains to be seen. In any event it seems clear that from age twenty-one personal demoralization increases with age among men.

Within each age group the men living in high economic status neighborhoods maintain the lower mean anomie scores. In each age group this difference is significant beyond the .001 level. Thus, differences in the age distributions of the neighborhoods do not explain the relationships between anomie and the economic status of the neighborhood.

There are differences between the neighborhoods with respect to religious

<sup>6</sup> Insignificant changes in the wording of items 3 and 5 were made. Compare these with Srole's wordings (13).

preference and frequency of attendance at religious services. The Protestants, and especially the Jews, for example, are overrepresented in the high economic status neighborhoods while the Catholics are overrepresented in the low economic status neighborhoods. Thus, differences in anomie between these religious groups, as might be predicted from Durkheim's work, may obscure or confound the differences in anomie between the low and high economic status neighborhoods. Comparisons made between the major types of religious preferences—none, Protestant, Catholic, and Jewish—show no significant differences in anomie with neighborhood economic status held constant. Within each religious group, excepting the Jews, however, men living in St. Francis Wood or Pacific Heights have significantly lower anomie scores on the average than men living in Outer Mission or Mission. In the case of the Jews the small number, four, living in the low economic status neighborhoods did not permit a meaningful comparison of anomie by neighborhood economic status within this religious group.

Similar findings are achieved when relating frequency of attendance at regular religious services to anomie. Holding neighborhood economic status constant, no differences in anomie appear between frequent and infrequent church attenders. However, significant differences in anomie still persist between the low and high economic status neighborhoods.

In Table 6 mean anomie scores are given by individual and neighborhood economic status. There are, of course, some men of low economic status living in Pacific Heights and St. Francis Wood, as there are a few men living in Mission and Outer Mission who can be classified as having high economic status themselves based upon their own occupation, education, and income. It can be seen from Table 6, as predicted from Merton's

TABLE 6  
*Mean Anomie Scale Scores by Individual Economic Status and Neighborhood Economic Status*

Individual Economic Status	Neighborhood Economic Status	
	Low (Mission and Outer Mission)	High (Pacific Hts. & St. Francis Wood)
High	3.4	2.7
Low	4.6	3.6

Significant differences:

Low vs. high individual economic status at low neighborhood economic status.\*

Low vs. high individual economic status at high neighborhood economic status.\*\*

Low vs. high neighborhood economic status at low individual economic status.\*\*

\*  $p < .05$ .

\*\*  $p < .001$ .

hypothesis, that low economic status men have higher anomie scores on the average than high economic status men. However, holding individual economic status constant, men living in high economic status neighborhoods still have lower anomie scores than men living in low economic status neighborhoods. Although the mean differences in anomie scores between the neighborhoods have been reduced by introducing individual measures of economic status, this difference is significant at the .001 level among the low economic status men and just barely misses significance at the .05 level among the high economic status men.<sup>7</sup> This suggests that the economic status of the neighborhood population as a unit may be related to anomie not merely because it reflects individual economic status, but also because it is an important variable in its own right. The question is, how does a neighborhood characteristic reflect or affect individual behavior? Two general notions can be found in the literature. The first considers the neighborhood as a selective variable; this refers to the process by which the social character of a neighborhood is perceived and evaluated by the individual and then rejected or selected as a place to live. The second considers the neighborhood as a conditioning variable; this refers to the effect of the neighborhood on the individual after he moves into the neighborhood. In the second case the higher anomie scores among the men living in the low economic status neighborhoods may reflect the fact that both the physical features and the general sub-cultural atmosphere are more favorable to an optimistic view of the world in the high rather than the low economic status neighborhoods.

#### DISCUSSION

It seems clear from the results of this study that anomie is inversely related to economic status. This is true whether economic status is measured by individual or neighborhood variables. This is consistent with Srole's findings and supports Merton's contention that differential access to economic success goals combined with a generally uniform expectation for economic success will result in anomie among those persons with the least opportunity to achieve such success.

Anomie is also related to social isolation, men who are relatively isolated having higher anomie scores on the average than men who are not so isolated. The wide variation in frequency of participation in informal and formal groups among urbanites is thus accompanied by some variation in the attitudes and sentiments of urbanites concerning their lot in the social order. It would be a distortion of the facts to characterize urbanites as

<sup>7</sup> Differences between the tract populations with respect to anomie in general still exist when individual occupation, income, and education are controlled separately.



being anomie generally, as some discussions of urban versus rural personality types imply. Rather it would be more realistic to refer to the wide variability in personal demoralization among urbanites as related to differences in frequency of informal and formal group participation, age, and position in the class structure.

The relationship between neighborhood economic status and anomie suggests that the economic character of the neighborhood population as a unit may play an important part in sorting out persons having different degrees of anomie. It also may directly affect the personal morale of the residents. This assertion is made more plausible by the fact that this relationship persists even when certain other variables are controlled (informal and formal group participation, rural-urban background, age, religious preference, frequency of church attendance, and *individual economic status*).

In conclusion, I would agree with Roberts and Rokeach that Srole has isolated an important variable with his anomie scale. In the final analysis the ultimate test of the theoretical considerations underlying the construction of any index, thus the ultimate test of the index's validity, is its predictive (and manipulative) utility. The weight of these findings suggests the verification of most of the hypotheses of this study; therefore the findings also constitute, to that extent, a partial validation of Srole's anomie scale.<sup>8</sup>

*Manuscript received: July 18, 1956*

*Revised manuscript received: December 10, 1956*

Wendell Bell

Department of Sociology

Northwestern University

Evanston, Illinois

#### REFERENCES

1. Bell, W., "The Utility of the Shevky Typology for the Design of Urban Sub-Area Field Studies," *Journal of Social Psychology*, in press.
2. Bell, W., and M. T. Force, "Social Structure and Participation in Different Types of Formal Associations," *Social Forces*, 1956, 34, 345-350.
3. Bell, W., and M. T. Force, "Urban Neighborhood Types and Participation in Formal Associations," *American Sociological Review*, 1956, 21, 25-34.
4. Bell, W., and M. D. Boat, "Urban Neighborhoods and Informal Social Relations," *American Journal of Sociology*, 1957, 62, 391-398.
5. BOWERS, R. V., "Discussion: An Analysis of the Problem of Validity," *American Sociological Review*, 1936, 1, 69-74.

---

<sup>8</sup> For a discussion of validity as logical consistency within a conceptual matrix see Bowers (5). Of course, if the hypotheses had not been verified then the question would remain as to whether Srole's scale were invalid or the hypotheses were, in fact, false.

6. MERTON, R. K. "Social Structure and Anomie," *Social Theory and Social Structure*, Glencoe, Illinois: The Free Press, 1949.
7. Roberts, A. H., and M. Rokeach, "Anomie, Authoritarianism, and Prejudice: A Replication," *American Journal of Sociology*, 1956, 61, 355-358.
8. Rokeach, M., "Rejoinder," a letter to the editor in the *American Journal of Sociology*, 1956, 62, 67.
9. Shevky, E., and M. Williams, *The Social Areas of Los Angeles*, Berkeley and Los Angeles, California: University of California Press, 1949.
10. Shevky, E., and W. Bell, *Social Area Analysis*, Stanford, California: Stanford University Press, 1955.
11. Srole, L., "Anomie, Authoritarianism, and Prejudice," a letter to the editor in the *American Journal of Sociology*, 1956, 62, 63-67.
12. Srole, L., "Social Dysfunction, Personality, and Social Distance Attitudes," Unpublished paper read at the annual meeting of the American Sociological Society, 1951.
13. Srole, L., "Social Integration and Certain Corollaries: An Exploratory Study," *American Sociological Review*, 1956, 21, 709-716.
14. Williams, R. M., et al, *Friendship and Social Values in a Suburban Community*, a working report from the Summer Seminar in Behavioral Science Research at the University of Oregon, mimeographed, 1956.

# **An Index of Marital Integration<sup>1</sup>**

BERNARD FARBER, *University of Illinois*

This paper is concerned with describing and testing an index of marital integration which aims at the following: (a) to pertain primarily to matters of consensus and interpersonal relations, (b) to require little time and skill to administer and score, (c) to refrain from asking the respondent to evaluate his marriage overtly or describe difficulties in the marriage, and (d) to be derived from explicit assumptions.

The index of marital integration is described, first, on the basis of assumptions involved and, secondly, as a procedure. The results of the application of the index to a sample of 99 married couples are then presented. Scores on the index are compared with social and emotional aspects of family interaction (10) and with perceived similarity between family members. The index scores are compared with interviewers' evaluations. In addition, scores on the index are related to personal adjustment in marriage.

## **ASSUMPTIONS UNDERLYING INDEX**

Studies of reactions of the family to the depression offer a variety of conceptual schemes for describing family integration in the face of crisis. Of these schemes, the one developed by Merton (9) seems to offer the simplest and most clear-cut analysis of integration. In his restudy of the Angell documents concerning the effects of the depression on the family, Merton tests three diagnostic concepts, namely, degree of social integration, character of interpersonal relationships, and level of pecuniary aspirations. Level of pecuniary aspirations refers to the content of a value which was related peculiarly to the depression and not to the social structure of the family itself. Therefore, only his first two concepts will be considered in the discussion of family integration. Since the categories used by Merton were bound closely to the specific population he was studying, some modification will be necessary in the present study.

It is assumed that family integration as a value system is based on the consensus of its members as to the rank-ordering of ends (11, cf. p. 203). The continued existence of the family as a group depends upon the successful performance of certain tasks such as socialization of parents and children, maintenance of a domicile, and economic activities. Successful

<sup>1</sup> The index described in this paper was developed for the study, "Effects of a Severely Retarded Child on Family Integration." This study is supported by a grant from the Mental Health Fund of the Illinois Department of Public Welfare to the Institute for Research on Exceptional Children, University of Illinois.

performance of these tasks can be viewed as ends in family life. Because of time limitations, extrafamilial commitments, and variations in systems of belief, the family members must decide which tasks should be given priority over others. These ends are thereby evaluated in decision-making and establishing routines and are ranked in order of preference. They describe the potentialities and tendencies of future action.

If the family with complete consensus did exist, it would be one in which all values were similarly rank-ordered by all its members. For any given decision, alternative means might be questioned or devised, but the ends would be agreed upon. Within the family, there would be mutual support of perspectives. The community of values then would provide a strong sense of identity among the family members. In addition, the family members would not feel threatened in decision-making situations. Thus, a common ranking of values, whether present initially or developed during the life of the group, provides a situation both congenial to effective mutual coordination of the life careers of the family members and compatible with effective socialization of the family members.

Commitment to a given ranking of values is not the same as commitment to roles within a particular group. Two individuals can agree on a given ranking of ends and yet each might be unwilling to commit himself to roles which the other expects of him to attain these values. The reverse may also be true. Two individuals can assume roles which the other expects, but might do this to attain different ends.

Sometimes, whether or not there is a high degree of community in ranking domestic values, family members fail to coordinate their roles effectively. The system of roles can then be said to be in a state of tension.

In the communication which accompanies the role tension, the self and spouse are redefined in terms of the character of interaction. Tempers may flare, arguments may occur, affectionate demonstrations may cease, decisions may be imposed, the actors may become sullen or avoid conversation. As these behaviors become the expected instead of the exceptional action in the association, they tend to be attributed to the other person or actor as part of the self. Conformity to role expectations is then considered superficial and difficult.

It is assumed that in the highly integrated family (a) the individual members are able to develop domestic and community roles while maintaining a sense of personal integrity, and (b) the family members are competent to meet crises without loss of commitment to one another and with a minimum of disruption of their domestic careers. These characteristics are inferential and should be tested by further research.

In summary, family integration is regarded as having two aspects. These

are integration of ends, which is defined as consensus on the rank-ordering of values by the family members, and integration of means, which for the purpose of the construction of the index is restricted to the mutual coordination of domestic roles. The degree of family integration then is determined by the extent to which family members agree on the rank-ordering of values and the degree of appropriate coordination of domestic roles. Failure to coordinate roles places the role system in a state of tension. Interaction of the family members under conditions of high role tension affects perceptions of one another by the family members.

A marital integration index appropriate to the conceptual scheme described above seems to be one which would enable (a) comparison of ranking by husband and wife on domestic values, and (b) estimation of role coordination on the basis of perceptions by husband and wife of one another.

#### COMPOSITION OF THE INDEX

In this section, the index of consensus will be described first; second, the index of marital role tension will be discussed; and, finally, the procedure for combining the two parts into a single estimate of marital integration will be presented.

##### *Index of Consensus*

In the construction of the consensus index, the aims were: (a) to stress those values which are relevant to decisions made in the family, (b) to permit the designation of specific hierarchies of domestic values, (c) to make definitions general enough so that the index would be applicable to a wide range of respondents, and (d) to take into account as many basic issues in the family as possible without becoming too lengthy.

The consensus index was administered in the following way:

Husband and wife were interviewed simultaneously in separate rooms to prevent collusion. A list of ten domestic values to be ranked in order of decreasing importance to family success was included in a questionnaire relating to family and community (8, cf. pp. 408-409). The respondent marked the rank of each item on the questionnaire. Occasionally, when the respondent's reading ability was found to be poor, the list of values was read to him or a paired comparison technique was used. The list of domestic values, as it was presented to the respondent, together with instructions for ranking, appears as Figure 1.

To measure the extent of agreement between the husband's and wife's rankings of domestic values, the Spearman rank correlation coefficient,  $\rho$ , is used. The size of the rank correlation coefficient is used as an index of the degree of consensus.

FIGURE 1

*Success in Marriage and Family*

Below are listed standards by which family success has been measured. Look through the list and mark 1 after the item you consider most important in judging the success of families (in the column headed *Rank*). Look through the list again and mark 2 after the item you consider next important. Keep doing this until you have a number after each item.

There is no order of items which is correct; the order you choose is correct for you. Remember, there can be only one item marked 1, one item marked 2, one item marked 3, . . . one item marked 10.

	<i>Rank</i>
<i>A place in the community.</i> The ability of a family to give its members a respected place in the community and to make them good citizens (not criminals or undesirable people).....	_____
<i>Healthy and happy children.</i> .....	_____
<i>Companionship.</i> The family members feeling comfortable with each other and being able to get along together.....	_____
<i>Personality development.</i> Continued increase in family members' ability to understand and get along with people and to accept responsibility.....	_____
<i>Satisfaction in affection shown.</i> Satisfaction of family members with amount of affection shown and of the husband and wife in their sex life.....	_____
<i>Economic security.</i> Being sure that the family will be able to keep up or improve its standard of living.....	_____
<i>Emotional security.</i> Feeling that the members of the family really need each other emotionally and trust each other fully.....	_____
<i>Moral and religious unity.</i> Trying to live a family life according to religious and moral principles and teachings.....	_____
<i>Everyday interest.</i> Interesting day-to-day activities having to do with house and family which keep family life from being boring.....	_____
<i>A home.</i> Having a place where the family members feel they belong, where they feel at ease, and where other people do not interfere in their lives.....	_____

*Index of Role Tension*

The index of role tension is composed of ratings by the respondents for both self and spouse on ten personality traits (6, p. 598). As reported by Farber and Blackman, the items were selected on the basis of a factor analysis on a sample other than that described in this paper. This was done to avoid using the same sample to select the items and test the index. The ratings and weights assigned to these ratings are: very much (-2), considerably (-1), somewhat (0), a little (+1), and hasn't the trait at all (+2). The total of the husband's ratings for himself and his spouse and the wife's ratings for herself and her spouse are used as the index of marital role tension. A high score is regarded as indicating high integration (i.e., low marital role tension) and a low score as denoting low integration (i.e., high role tension).

Although only ten personality-trait items are scored in the index, the list of traits presented to the respondent included several "positive" or

"neutral" traits. In the list below, the items with an asterisk are used in the scoring. The order of the sixteen traits rated by the respondent is: sense of humor, sense of duty, stubborn,\* gets angry easily,\* feelings easily hurt,\* nervous or irritable,\* easygoing, moody,\* jealous,\* likes to take responsibility, dominating or bossy,\* critical of others, easily excited,\* shy, likes belonging to organizations, easily depressed,\* and self-centered.\*

#### *Combining Consensus and Role-Tension Indices*

While, for analysis, means and ends are treated as separate entities, the ranking of values is not wholly independent of the coordination of roles. For example, while both the husband and wife may rank "healthy and happy children" first, they may differ on succeeding ranks. The husband may follow the children ranking with "religious and moral unity," "economic security," and "a place in the community." The wife, however, may follow the ranking on children with "emotional security," "personality development," and "companionship." It seems reasonable to imagine that the difference between husband and wife in these other rankings would influence their interpreting and handling the problems of their children and possibly generate resentment or argument. At the time the couple completes the integration form, the difference in handling the children may not yet have been crystallized in interpersonal perceptions. This strain in the system of roles, potential or actual, however, will be accounted for in the consensus section of the index.

The index of marital integration was formed by arbitrarily adding ordinal scores on the consensus and marital-role tension indices. Since at present there is no evidence as to which is the more important, both indices are given equal weight. Future testing of the integration index on several criteria relating to mental health and the ability of the family to weather crises should yield a more precise scheme for weighting and combining the two indices.

In order to assign tentative weights, data on consensus and role tension for the first 200 cases to be interviewed in the study for which this index was developed were arranged in quartiles. The lowest quartile received a 0 weight, and the top quartile a weight of 3 for each of the two indices. The weights for the two indices were added to provide the index of marital integration. A case falling into the top quartile for both consensus and role-tension indices received a marital integration score of 6; a case in the lowest quartile for both indices received a score of 0.

#### CORRELATES OF THE INDEX

The extent to which the index accurately summarizes marital integration will be examined in several ways: (a) consistency with a current theory



and/or knowledge, (b) consistency with personal experience and impressions, and (c) consistency with past indices of marital success.

#### *Consistency with Current Theory and Knowledge*

The consistency of the index of marital integration with theory and findings in two areas were examined: (a) the social-emotional aspects of interaction and (b) perceived similarity of family members.

The particular hypotheses tested in the study were developed in an attempt by the writer to relate the conclusions of his previous research on courtship and the early years of marriage (5; 3, pp. 453-454, 587; cf. 4) and experience with families having mentally retarded children to the work of Parsons and Bales (10).

For this part of the study, the index of marital integration was applied to 99 white families in Chicago and an industrial city in eastern Illinois. In all these families there was at least one child of normal intelligence, aged five to eighteen, living at home and attending school.

The sample in eastern Illinois consisted of 54 couples who volunteered in response to a request by the local PTA council. The 45 couples in Chicago were reached through the cooperation of the associations formed to promote the welfare of the mentally retarded. All those from Chicago were parents of a child in a state institution for the mentally deficient. The interviewing of these samples was done from October, 1955, to April, 1956.

Families in which any children had been born in a previous marriage or in which there was a mentally retarded child living at home were not used in the present analysis.

Sample bias will not be discussed in this paper but will be described in the general report of the study.

Of the families in the sample used, the mean age was 41.0 for men and 38.1 for women. The mean years of schooling completed was 12.4 for men and 11.8 for women. The couples had been married a mean of 16.1 years and had a mean of 3.1 children. On the whole, the sample can be described as middle class, but not especially biased toward upper middle class.

Inasmuch as participation in the study was on a volunteer basis, tests of statistical significance based on random selection have only limited validity in assessing results. The reader, therefore, should base his evaluation of these results not only on the statistical tests but also on the consistency of the various findings with one another. In the statistical tests, the .05 level of confidence was used as the criterion for rejecting null hypotheses.

#### *Marital Integration and Social-emotional Aspects of Interaction*

On the basis of a specialization of task and social-emotional functions in the family, Parsons and Bales (11) assign the task-oriented role to the

husband and social-emotional role to the wife. They regard this role differentiation as subordinate to the common value system in the family. In terms of a means-ends schema, the marriage lacks integration to the degree that this differentiation of roles (or means) takes precedence over the common value system (or ends).

Whether the common value system is oriented toward tasks or toward social-emotional aspects of interaction may have consequences for the degree of marital integration. Because the woman's role as wife and mother revolves around the home, it seems that there is generally less opportunity for variation in her general orientation than in her husband's (3, p. 587). That the husband's orientation toward social-emotional aspects of interaction is more important than the wife's in achieving marital success is supported in a study by Chusid, Haydon, Manniche, and Smith (4). Applying partial correlation to a rating scale for use with case material, they found "companionship" or intimate interaction in engagement an accurate predictor of marriage adjustment for men, but not for women. Thus, the husband who agrees with his wife on the importance of social-emotional aspects of interaction tends to promote general marital integration.

If the index is a valid measure of marital integration, hypotheses relating integration to social-emotional aspects of interaction as opposed to task-oriented aspects of interaction should be supported in the study of the 99 married couples. The hypotheses tested were:

Hypothesis 1. *The degree of marital integration varies directly with the ranks assigned by the husband to domestic values pertaining to social-emotional aspects of interaction.*

Hypothesis 2. *The ranking of items relating to social-emotional aspects of interaction by wives tends to be higher than the ranking by their husbands.*

In the analysis, those values described in Figure 1 which seemed to be related to the development of a system of personal relations were regarded as social-emotional and all of these were assigned a rank of 3. The social-emotional values were companionship, personality development, a home, emotional security, and satisfaction with affection shown. The remaining values, which seemed to describe the family in terms of its continuity, administration, and place among other institutions, were regarded as task-oriented. The values regarded as task-oriented were a place in the community, everyday interest, healthy and happy children, moral and religious unity, and economic security. The task-oriented values were given a rank of 8. Thus, the higher the rank correlation between the ideal ranking just described and the ranking by an individual, the more can the individual be said to emphasize social-emotional interaction in his family life.

The ranking of values by husband and wife enters into the computation of both the consensus index and the index of social-emotional valuation. Because the ideal ranking of social-emotional valuation is composed of two tied

ranks (with five items for each rank), however, the husband and wife can have the same score on the social-emotional value index and widely differing scores on the consensus index. In the weighting of consensus scores for the marital integration index,  $-.01$  or lower (the lowest quartile) is given a weight of 0 and  $+.50$  or higher (top quartile) is given a weight of 3. Eighty-six per cent of the social-emotional value scores fell between  $-.38$  and  $+.45$ . If the husband's social-emotional value score was  $-.38$  and his wife's  $+.45$ , the possible range of their consensus scores would be  $-.99$  to  $+.60$ , a sufficient range to place them in any quartile of consensus scores. In all but three cases, a consensus score considerably higher than  $+.50$  and lower than  $-.01$  was possible.

For the analysis between husband's social-emotional valuation and marital integration, the three cases whose consensus scores were restricted by the social-emotional value scores too severely to fall into any quartile were not used. Table 1 shows the relationship between the husband's social-emotional orientation and marital integration. The chi-square of 7.754, with two degrees of freedom, is significant at the .05 level.

A second test of Hypothesis 1 was made. By induction, it had been found that as the social-emotional value rho increases from .00 to  $\pm 1.00$ , the range of possible consensus scores decreases. For the second test of Hypothesis 1, therefore, to permit the widest possible range of consensus scores, only those women's social-emotional value scores in the range of  $+.17$  to  $-.17$  were used in this analysis. When both husband and wife have a social-emotional value score of  $+.17$ , the possible range of their consensus scores is from  $+1.00$  to  $-.92$ . There were 43 cases in which the wife's social-emotional value score fell in the range  $+.17$  to  $-.17$ . The range of husband's social-emotional scores was  $+.87$  to  $-.45$ , with a median at  $+.03$ . A Spearman rank correlation was computed for these 43 cases between the man's social-emotional value score and the marital integration score for the couple. The resulting rho was  $+.29$ . When a one-tailed t-test was applied ( $t = 1.918$ , 41 degrees of freedom), the result was significant at the .05 level (12, p. 207).

TABLE 1  
*Marital Integration and Husband's Social-Emotional Valuation*

Husband's Social-Emotional Valuation (Rho)	Marital Integration			Total
	0-2	3-4	5-6	
.10 or higher .....	4	15	26	45
.09 or lower .....	15	18	18	51
Total .....	19	33	44	96

Note:  $\chi^2 = 7.754$ . Significant at .05 level

To test the hypothesis that the woman's social-emotional value score tends to be higher than her husband's, the Wilcoxon matched-pairs signed-ranks test was applied (12, pp. 75-83). The Wilcoxon test takes account not only of the direction of difference but also of the size of the difference. With the husband having exactly the same social-emotional value score as his wife in four cases, the number of cases used in this test was 95. With a critical ratio of 3.32 and the probability inferred on the basis of a one-tailed test and a normal distribution, the results are significant at the .0005 level.

Hypotheses 1 and 2, therefore, are supported by the data.

#### *Perceived Similarity and Marital Integration*

Studies of assumed similarity or personal identification in small groups have found that the degree to which individual A perceives personal attributes of individual B as similar to his own varies inversely with the degree of social distance between A and B as seen by A (7). It has been indicated above that marital integration is associated empirically with high valuation placed on social-emotional aspects of interaction. If we assume that valuation of social-emotional aspects of family life is related to intimacy in social relations (or lack of social distance), perceived similarity should also vary directly with marital integration (3, pp. 453-454).

The organization of roles in the family should be taken into account in assessing the relationship between integration and perceived similarity. From the Parsons and Bales discussion cited above, it seems that the wife acts in a mediating role, with the husband and children at the extremes. The wife is then in the position of balancing the husband's demands and expectations with the children's. The husband who stresses social and emotional aspects of interaction supports the wife's role in the socialization of the children (which she tends to regard as her primary task in the family) and enables her to develop a close relationship with the children. Therefore, while in a highly integrated marriage it is important for the husband to identify with the wife but not necessarily with the children, the wife should identify especially with the children. Since the husband in the integrated marriage tends to stress social-emotional aspects of family life, the wife is not then caught in a situation of conflicting expectations. With perceived similarity as an index of social intimacy, the hypotheses tested were:

Hypothesis 3. *For husbands, the similarity between ratings of self and spouse tends to vary directly with the marital integration index score.*

Hypothesis 4. *For wives, the similarity between ratings on self and children tends to vary directly with the marital integration index score.*

Conversely, the relationship between marital integration index and husband's ratings of self and children and wife's ratings of self and husband

should be in the range of chance expectation. (A positive result beyond chance expectation, however, would not vitiate the theory.)

The procedure for classifying relationships as high perceived similarity and low perceived similarity involved the ratings on the ten personality traits used in the marital integration index. If the respondent rated himself and the individual with whom he was comparing himself identically for six or more of the ten traits, the relationship was classified as high perceived similarity. Five or fewer identical ratings for the other individual and himself placed the relationship in the low perceived similarity category.

Ratings for children aged two or under and for retarded children frequently were not considered as valid by parents who provided ratings to fulfill the formal requirements of the interview and, therefore, were not included in the analysis.

Another adjustment made in the data was to remove those families with five or more children for this analysis. There were eleven such families. This was done to eliminate the possibility of influencing the data where the large numbers themselves may be a factor in perceived similarity. The personality ratings in three other families showed many omissions especially for the children. These families were also omitted from the analysis. Eighty-five families remained in the study of perceived similarity.

The relationship between marital integration scores and the husband's perceived similarity between his wife and himself is shown in Table 2. The chi-square of 7.315, with two degrees of freedom, is significant at the .05 level.

The data on marital integration scores and the similarity between ratings by wives of themselves and their children are reported in Table 3. With a chi-square of 8.443, with two degrees of freedom, the probability is less than .05 that chance alone is responsible for the relationship between the perceived similarity and marital integration scores.

Contingency tables in which the chi-square did not attain the size required for the .05 level of significance were on the relationship between

TABLE 2  
*Marital Integration and Husband's Perceived Similarity between his Wife and Himself*

Perceived Similarity	Marital Integration			Total
	0-2	3-4	5-6	
High.....	4	14	22	40
Low.....	14	17	14	45
Total.....	18	31	36	85

Note:  $\chi^2 = 7.315$ . Significant at .05 level

TABLE 3

*Marital Integration and Wife's Perceived Similarity between at Least One of Her Children and Herself*

Perceived Similarity	Marital Integration			Total
	0-2	3-4	5-6	
High for at least one child.....	3	14	21	38
Low for all children.....	15	17	15	47
Total.....	18	31	36	85

Note:  $\chi^2 = 8.443$ . Significant at the .02 level

marital integration scores and the husband's ratings of self and children and the wife's perceived similarity of her husband and herself.

The results on perceived similarity, thus, were consistent with Hypotheses 3 and 4 and support the proposition that the index is a valid measure of marital integration.

There is no necessary association between the personality trait score for the marital integration index and the perceived similarity classification. If there were, the wife's perceived similarity would also be related, beyond chance expectation, to marital integration. Nor were there found any "perfect scores" in the marital integration scoring of the personality traits to create a halo bias.

#### *Consistency with Personal Experience*

Interviewers' reports will be presented below to convey nuances in the correspondence between the index and the situation it is supposed to describe.

Each interviewer for the study was asked to give his general impressions concerning each interview. The oral section of the interview lasted from one-half hour to two hours. This section was concerned mainly with parents' conceptions of their children. All the interviewers had had previous interviewing experience and were either social workers or advanced graduate students in the social sciences. Inasmuch as marital integration scores were not computed until the interview reports were placed in the files (and the interviewers did not know how to compute these scores), the interviewer could not base his discussion of the interview upon the integration score. In addition, a given interviewer interviewed the husband or the wife, but not both. There were no restrictions on how these impressions were to be reported; the interviewers were merely asked to present what they considered important for interpreting the interview. These interviewer impressions are being analyzed at present and will be reported elsewhere.



To illustrate a correspondence between integration score and marital situation, interviewer impressions will be presented below for two cases, selected at random, of couples with extreme scores. These couples did not have a mentally retarded child.

Case A. Marital integration score: 6.

*Impressions of wife's interviewer:* The wife is as cooperative at home with her husband as she is in community affairs. She takes care of the bookkeeping involved in her husband's job. . . . The outstanding quality of this woman is her friendliness. She apparently has good relations with her in-laws. . . . I would conclude that this is a relatively well-integrated family unit. The marital relationship is especially well integrated.

*Impressions of husband's interviewer:* I sort of had a hunch that there must be some area in this man's life undisclosed to me that was where he was really emotionally involved and identified. But then again, it may very well be that he was just a big, easy-going, placid, lethargic man who plugged along without much involvement with anything beyond the biological fundamentals. All in all, this family seems to be getting along okay. I got the feeling that more was going on in this family than I found out about. But from what I did see, it seems to be a fairly phlegmatic family going along without too many high points and presumably not too many low points either, just rocking along. I presume they will continue to do so. I don't see any reasons why they should not.

Case B. Marital integration score: 0.

*Impressions of wife's interviewer:* The woman emphasized she didn't know the income of her husband, explaining that he gives her an allowance on which she carries out her duties as housewife. In a low-toned voice, she said she didn't know what the total income of her husband was.

*Impressions of husband's interviewer:* The major content of this man's life is his work, occupation, and his sons. He rarely spoke of his wife and when he did, it was with a most impersonal "her" or "she." He did not automatically include her in his thoughts about himself and his life. I got the feeling that the wife was a necessary appendage to the family, a very useful member, but nevertheless, more or less ignored. . . . In summation, I would say that the central things in this man's life were his work and his sons. . . . The trio of males seem to comprise the solidary family unit and the wife and mother is a necessary adjunct to the well functioning of these three. This is, of course, a picture gotten from interviewing only the husband and from talking, except when we first arrived, with only the three males of the family. . . . I suspect the husband has accepted the wife as she is and is no longer trying to change her or his family life in general.

An interviewer's impression of a more extreme case with a marital integration score of 0 is reported below:

When we had coffee together [after the interview], she [the wife] would ignore Mr. X completely or just give him rueful looks. She would seem annoyed if I directed some of the conversation to him as I did deliberately to see what their relationship was like. . . . She has a great deal of hostility toward her husband and a general dissatisfaction with her marriage.

Remarks by still another interviewer on the general demeanor of male respondents in interaction with him are presented below. The cases reported



are in a subsample of every sixth family in which one of the children is in an institution for the mentally retarded. All cases done by this interviewer in the subsample are reported. An assumption made in interpreting the following remarks by the interviewer is that the respondent's behavior may reflect his general approach to interpersonal relations and that perhaps the rapport in the interview itself is influenced at least in part by the respondent's ability to establish and maintain interpersonal relations (cf. 13, p. 176).

Case 1. Marital integration score: 6.

Respondent was very friendly and open with me throughout the interview.

Case 2. Marital integration score: 6.

Respondent was not really verbal. . . . He was, however, very friendly and open with me.

Case 3. Marital integration score: 4.

I felt he answered thoughtfully . . . rather than with thought of evasion.

Case 4. Marital integration score: 4.

There was a tendency to evade some questions. . . . [He had] an excellent sense of humor.

Case 5. Marital integration score: 3.

In the beginning, respondent said he was reluctant to be interviewed. . . . After I explained a little to him, he seemed less against it. . . . He became quite verbal after a while and gave direct answers to all questions.

Case 6. Marital integration score: 3.

He had a tendency to begin discussing his ulcers. . . . When asked a question, I felt he answered honestly, but many times I had to rephrase and repeat to get more than just short answers from him.

Case 7. Marital integration score: 2.

Respondent seemed quite disorganized in his thinking and answering questions. I don't know whether this was evasiveness or just lack of ability to organize his thoughts.

The above impressions by the interviewer can, of course, lead to the interpretation that the score on the marital integration index is strongly influenced by the rapport established. This interpretation is rejected here because the spouse, in all cases, was interviewed by another interviewer and the responses of both husband and wife were used in computing the integration index.

In this section, interviewers' impressions have been presented to indicate some correspondence between the marital integration score and other observations of the family situation.

## MARITAL INTEGRATION AND PERSONAL ADJUSTMENT IN MARRIAGE

To provide comparability and continuity with studies using the adjustment criterion, the relationship between the index of marital integration and two indices of personal adjustment was tested.

While in some ways marital integration and personal adjustment may be regarded as similar, in the light of the assumptions made in constructing the marital integration index there are important differences. One distinction between the personal adjustment and integration concepts is that the husband and wife who seek different ends but who evolve mutually accommodating roles may be termed as personally adjusted, but their association in marriage is not highly integrated. A second difference is that personal adjustment is essentially a psychological construct, its focus being the individual's happiness or regret, his feelings of satisfaction, and his sense of both sociability and lack of conflict in the marriage. Integration in marriage, however, rests on actual agreement in value hierarchies and the development of roles appropriate to these value hierarchies. Generally, marital integration seems more descriptive of consensus in interaction than does personal adjustment.

The two indices of personal adjustment used in evaluating the integration index were (a) extent of happiness reported and (b) frequency of regretting marriage. Past study has shown these two marital adjustment indices to be highly intercorrelated (2, p. 66). In past usage, however, when happiness and regretting marriage were applied to assess marital adjustment, the result was interpreted only for the individual. For example, in the Burgess and Cottrell and the Terman studies, it was possible for one member of the couple to be very happy and have a high adjustment score while his spouse was very unhappy and had a low adjustment score. In the present study, the happiness ratings for the *couple* were compared with that couple's marital integration score. One index (happiness) was used to present a positive statement of their marital adjustment and the other (regret marriage) to present a negative statement of their adjustment.

In completing their written forms, the individuals answered the following questions:

Do you ever regret your marriage: (check): Often —; sometimes —; a few times —; never —.

Do you think your mate ever regrets having married you? (check): Often —; sometimes —; a few times —; never —.

Everything considered, how happy has your marriage been for you? (check) Very happy —; happy —; somewhat happy —; average —; somewhat unhappy —; unhappy —; very unhappy —.

Everything considered, how happy has your marriage been for your mate? (check): Very happy —; . . . very unhappy —.

The questions about the mate were asked to give the respondent an opportunity to record his own projections and insecurity or attitudes which his spouse may fail to report.

The couples were classified, first, on whether the husband or wife reported that either one had ever regretted his marriage and, second, on whether either one had rated himself or his spouse as anything other than "very happy" or "happy." The "never regretted marriage" category included only those cases in which both the husband and wife reported that neither had ever regretted the marriage. Similarly, the "happy couples" category included only the cases in which the husband and wife reported ratings of "very happy" or "happy" both for himself and his spouse.

In the 99-case sample, for both indices, there was a positive relationship between personal adjustment and the marital integration score. The results, however, did not attain statistical significance.

In order to check the possibility of a Type II error (i.e., the null hypothesis is accepted when it is, in fact, false), it was decided to test the relationship between the same two indices of personal adjustment and the marital integration score on the families with a mentally retarded child living at home. This test had already been planned as part of the research project, regardless of the results on personal adjustment in the 99-case sample.

The interviewing of couples with a severely mentally retarded child living at home was completed in October, 1956. These couples, like those with a retarded child in an institution, lived in or near Chicago and had been contacted through parents' organizations for promoting the welfare of the mentally retarded. In this sample, the mean age was 41.3 for

TABLE 4  
*Marital Integration Scores and Reports by Both Husband and Wife that Neither Has Ever Regretted the Marriage*  
(Families with a Mentally Retarded Child at Home)

Marital Integration Score	Marriage Never Regretted	All Other Reports	Total
6	14	9	23
5	18	12	30
4	18	27	45
3	12	29	41
2	6	20	26
0, 1	4	17	21
Total.....	72	114	186

Note:  $\chi^2 = 18.172$ . Significant at .01 level

TABLE 5

*Marital Integration Score and the Rating by Both Husband and Wife that Spouse and Self Are Happy or Very Happy*  
(Families with a Mentally Retarded Child at Home)

Marital Integration Score	Happy Ratings for Couple	All Other Ratings	Total
6	19	4	23
5	22	8	30
4	28	17	45
3	26	15	41
2	11	15	25
0, 1	8	13	21
Total.....	114	72	186

Note:  $\chi^2 = 15.133$ . Significant at .01 level

husbands and 39.2 for wives; the mean number of years of education was 12.2 for husbands and 11.5 for wives; they had been married a mean of 14.9 years. With Negro cases eliminated, there remained 186 married couples in the sample.

The results on never regretting marriage as related to the marital integration score are presented in Table 4. So that theoretical frequencies would not be smaller than five for computing the chi-square, integration index categories 0 and 1 were combined. With five degrees of freedom, the chi-square of 18.172 is significant at the .01 level.

The findings on happiness in marriage as compared with the marital integration index appear in Table 5. With five degrees of freedom, the chi-square of 15.133 is significant at the .01 level.

The existence of a positive relationship between the marital integration score and indices of personal adjustment is thus supported by the data of the 186-case sample.

#### DISCUSSION

In this paper, an attempt was made to describe an index of marital integration, along with assumptions underlying this index. Hypotheses were then developed and tested to determine whether family attributes derived from other observations and theoretical formulations are associated empirically with high and low marital integration as measured by the index.

Perceived similarity in small groups and the Parsons-Bales instrumental versus social-emotional dichotomy provided a framework for relating the index of marital integration to family attributes. It was found that, for white, urban, middle-class families, with at least one child of school age, as

measured by the index, marital integration tends to vary directly with:

1. The husband's emphasis on companionship or social-emotional ends in his family-value hierarchy.
2. The degree of identification (perceived similarity) of the husband with his wife.
3. The degree of identification of the wife with at least one of her children.
4. The personal adjustment of the husband and wife in marriage (in a second sample).

It was also found that, consistent with the Parsons and Bales discussion, women tended to rank values related to social-emotional aspects of interaction higher than did their husbands.

Remarks by interviewers were also introduced to convey some of the nuances of the correspondence between the integration index and the marital situation as observed.

As noted in several places in the paper, this report on the marital integration index will be supplemented by others offering additional substantiating data and possibly suggestions for administering and/or revising the index.

*Manuscript received: July 27, 1956*

*Revised manuscript received: January 7, 1957*

Bernard Farber  
Institute for Research on Exceptional Children  
University of Illinois  
Urbana, Illinois

#### REFERENCES

1. Burgess, E. W., and H. J. Locke, *The Family* (2d ed.), New York: American Book, 1953.
2. Burgess, E. W., and L. S. Cottrell, *Predicting Success or Failure in Marriage*, New York: Prentice-Hall, 1939.
3. Burgess, E. W., and P. Wallin, *Engagement and Marriage*, New York: J. B. Lippincott, 1953.
4. Chusid, R., E. M. Haydon, E. Manniche, and W. Smith, "The Revised Burgess and Wallin Rating Scale as an Instrument for Predicting and Measuring Integration in Marriage," M.A. thesis, University of Chicago, 1954.
5. Farber, B., "An Evaluation and Revision of the Burgess and Wallin Rating Scale as a Prediction Instrument," Ph.D. dissertation, University of Chicago, 1953.
6. Farber, B., and L. S. Blackman, "Marital Role Tensions and Number and Sex of Children," *American Sociological Review*, 1956, 21, 596-601.
7. Fiedler, F. E., "Assumed Similarity Measures as Predictors of Team Effectiveness," *Journal of Abnormal and Social Psychology*, 1954, 49, 381-388.
8. Kluckhohn, C., et al., "Values and Value-Orientations in the Theory of Action,"

- in T. Parsons and E. A. Shils (eds.), *Toward a General Theory of Action*, Cambridge, Massachusetts: Harvard University Press, 1951.
9. Merton, R. F., in E. W. Burgess, et al., *The Restudy of the Documents Analyzed by Angell in The Family Encounters the Depression*, (hctographed) New York: Social Science Research Council, 1942.
  10. Parsons, T., and R. F. Bales, *Family, Socialization and Interaction Process*, Glencoe, Illinois: The Free Press, 1955.
  11. Parsons, T., and E. A. Shils, "Values, Motives, and Systems of Action," in T. Parsons and E. A. Shils (eds.), *Toward a General Theory of Action*, Cambridge, Massachusetts: Harvard University Press, 1951.
  12. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill, 1956.
  13. Stanton, H., K. W. Back, and E. Litwak, "Role-playing in Survey Research," *American Journal of Sociology*, 1956, 62, 172-176.

# Stimulus Conditions as Factors in Social Change<sup>1</sup>

EDWIN COHEN, *Educational Research Corporation*

Much work has been done in the disciplines of sociology, anthropology, political science, history, and economics on the subject of social change. The purpose of this study is to present a laboratory prototype or paradigm of social change, testing the hypothesis that social change will result from a change in the stimulus situation confronting the group. The specific problem of this study was the nature of the social change, or norm shift, in terms of shifts in agreed judgments of a stimulus situation, which will result from the introduction of a new anchorage to the group. The hypothesis tested may be stated as follows:

When an anchorage well outside the original range of stimuli is introduced, group judgmental norms will shift, exhibiting contrast effect.<sup>2</sup>

## PROCEDURE

### *Background*

The present experiment was an extension of those studies concerned with the formation of judgmental scales by the method of single stimuli (3, 4, 7). Verbal material was used. There were, however, two important differences between this and previous studies.

1. In the present study, the anchors were not explicitly defined as such by the experimenter, but the subjects were free to choose for themselves which stimuli constituted the ends of the scale.

2. The present study did not deal merely with the formation of individual scales of judgment but with the formation of group norms concerning the placement of items on a scale by interacting pairs.

This experiment was designed to show that a group norm will be modified

<sup>1</sup> This paper constitutes a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oklahoma. The writer would like to thank Dr. Muzaffer Sherif, who directed the dissertation, for his invaluable aid.

<sup>2</sup> For those readers unfamiliar with the psychophysics of absolute judgment, a brief explanation might be in order. Suppose a subject, after being instructed to describe tones as low, medium, or high, were presented a series of tones scattered over the range 200 to 400 cycles per second. After a few trials, he would probably call a 300-cycle tone "medium." After a dozen or so, suppose that 2,000 cycle tones were interspersed among those in the 200-400 range. The subject would then likely call a 300-cycle tone "low," though he previously called it "medium." Thus his judgment of the 300-cycle tone shifted in a direction away from the 2,000-cycle tone, called an *anchor* since it serves to tie down one end of the scale. The phenomenon is called *contrast effect*, since by contrast with the 2,000-cycle tone the 300-cycle tone seems low.



to meet changed conditions. The group norms in this case were the agreed ratings by pairs of subjects, on a scale of undesirability, of 30 "moderately undesirable behaviors." The changed conditions were introduced by embedding these behaviors in the context of "very undesirable behaviors." The group norm should show what, on the individual level, has been termed contrast effect (2); the rating norms should shift away from the "very undesirable" end of the scale, where the anchor is introduced.

To demonstrate this shift in norm, it was necessary to confirm three sub-hypotheses:

Subhypothesis 1. A group norm was actually formed in the experimental situation. This will be shown by comparing the differences on separate, independent ratings by the two subjects of the same pair with the differences on ratings of subjects of different pairs. If a group norm is operative, two subjects from the same pair should show greater uniformity in their ratings when not in the physical presence of their "partners" than would two subjects not from the same pair. Formation of such group norms has been shown by many experimenters, starting with Sherif (5, 6). We must, however, show the existence of a norm before we can demonstrate its shift.

Subhypothesis 2. The 30 behaviors represented to be "moderately undesirable" were actually perceived as less undesirable than the "very undesirable" behaviors in which they are later embedded. This will be shown by comparing subjects' ratings of the behaviors placed by the experimenter in these two categories. This subhypothesis, too, is ancillary and is formally stated to sharpen the statement of the antecedent conditions for the main hypothesis, the "if" in the "if \_\_\_\_\_, then \_\_\_\_\_" statement.

Subhypothesis 3. The norms shifted, that is the 30 "moderately undesirable behaviors" are rated as less undesirable when embedded in the "very undesirable behaviors" than when not so embedded.

#### *General Procedure*

In order to secure data for these comparisons, pairs of subjects first jointly rated statements of moderately undesirable behavior (Mubs), e.g., "fishing without a license," on a graphic rating scale. Then subject pairs of the experimental sample rated these same statements along with additional statements of very undesirable behavior (Vubs), e.g., "kidnapping a baby for ransom," while pairs of the control sample rated the Mubs along with similar additional statements (Amubs), e.g., "hunting without a license." Lastly, each subject rated the same items presented in the second series (Mubs plus Vubs for experimental, Mubs plus Amubs for control) when not in the presence of his partner.

The use of a control sample made possible the statistical control of temporal or other extraneous factors which might have caused subjects to change their ratings from one presentation to another.

### *Subjects*

Twenty-five pairs of subjects were used for the experimental sample, and a like number for the control sample. The individuals in each pair were like-sexed and were unacquainted with each other prior to the experiment. All subjects were white summer-school students. Most were undergraduates, no subject had any course work in social psychology, and no subject correctly guessed the purpose of the experiment. In each sample, there were 5 female and 20 male pairs, a situation reflecting subject availability.

### *Apparatus*

A vertical screen with two horizontal, parallel slots, an inch and a half apart, was placed on a table between the experimenter and the pair of subjects, who sat side by side. This device will be referred to as the rating board. Each slot had in it a block with tongue affixed, so the block could be slid from one end of the slot to the other, or removed from the board entirely. Each block could be slid 300 millimeters. One block was colored red, the other black. Beyond each end of the corresponding slot, identifying colored squares, red and black respectively, were placed.

Behind each slot a 300 millimeter ruler was placed so that the position of a mark on the tongue of each block could be read to the nearest millimeter. These rulers were, of course, not visible to the subjects. Since the rulers were facing the experimenter the reading at the subjects' extreme *left* was 300, and extreme *right*, 0.

This apparatus is a kind of graphic scale; only ends of the scale furnish the subject with reference points. No other reference points are labeled, either by number (as in numerical scales), words (as in Likert-type scales), or lack of homogeneity in the apparatus which would allow a subject to differentiate points on the scale. A subject could make ratings differing by many millimeters without being aware of the discrepancy between the ratings. The blocks were removed from the apparatus after each rating to make it more difficult for subjects to compare a rating with the previous one.

A scale of this sort was found to be necessary when pretests, using a numerical scale, showed that subjects, having attached a number label to an item, were loath to change the label.

### *Stimulus Material*

For the experimental sample, the stimulus material consisted of 30 moderately undesirable behavior statements (Mubs) and 14 very undesirable behavior statements (Vubs). For the control sample, the 14 Vubs were replaced by 14 auxiliary moderately undesirable behavior statements (Amubs) which did not duplicate any of the 30 Mubs.

Most of the rating items were selected from a pool of 175 items, many adapted from McGarvey (3). Ten subjects, not used in the main part of the

experiment, rated these 175 items for undesirability on an 11-point scale; the Mubs and most Amubs items were chosen from those near the mid-point of the scale in average rating; the Vubs from those at the most undesirable end of the scale.

However, as was mentioned earlier, the justification for the classification of items (into Mubs, Amubs, and Vubs) was an empirical rather than a historical one, that is, the correctness of our classification was justified not on the basis of their origin, or how the ten preliminary subjects rated them, but on the basis of how the experimental and control samples used in the body of the experiment rated them.

The 30 Mubs were rated on graphic rating scales three times by each subject. The first rating was on the rating board, and for it the subjects were read the 30 Mubs in order.<sup>2</sup> On the second rating, also on the rating board, the 30 Mubs were embedded in the 14 Vubs (for the experimental sample) or the 14 Amubs (for the control sample). Embedding was done for the experimental sample by presenting five Vubs before the Mubs series, and placing a Vubs after each three items in the Mubs series. For the control sample, the Amubs items were used rather than the Vubs items.

The third rating, done independently by the subjects in separate rooms, was made on 300-millimeter graphic rating scales. The item list was the same as that used for the second rating; it had the 30 Mubs as well as the 14 embedding items.

### *Instructions to Subjects*

After being introduced to their "partners," each pair was given the following instructions:

"I've asked you to help with this study, which is designed to get at the opinions people reach after discussing certain topics. We will follow this procedure: I will read to you statements of things people sometimes do. All these things are usually considered wrong or undesirable, for various reasons, but not all to the same degree. After I read each statement, I want

<sup>2</sup> To save printing costs, the following have been deposited with the American Documentation Institute:

Appendix I: Mubs items given both experimental and control samples during presentation 1.

Appendix II: Rating board item sequence for the experimental sample on presentations 2 and 3.

Appendix III: Rating board item sequence for the control sample on presentations 2 and 3.

Appendix IV: Instructions for paper graphic rating scale.

Appendix V: Questionnaire filled out by all subjects.

For all these materials, order Document No. 5091 from American Documentation Institute, Library of Congress, Washington 25, D. C., remitting \$1.25 for 35-mm microfilm or \$1.25 for 6-by-8 photo copies. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

you to give your reaction to it by setting these little blocks along the scale. You, \_\_\_\_\_, work with the red block, and you, \_\_\_\_\_, work with the black one. Use the slot marked on each end with the color of your block. The more wrong or undesirable the behavior, in your opinion, the further to your right the block should be put.

"If the two of you do not place your blocks in the same position for a given statement, I would like you to discuss the behavior between you, and try to arrive at a rating which is mutually agreeable. Please tell me when you have reached agreement on the rating for an item, or when you feel that you cannot compromise the difference between you. When you give me the word, we'll go to the next item.

"When you rate the items, try to keep an objective point of view. That is, rate each item not on the basis of how wrong or undesirable it would be for you personally to do it, but on the basis of how wrong or undesirable you think it would be for other people to do it, without any reference to your own behavior. Try to be aware of fine differences in degrees of wrongness or undesirability; all these things are usually considered wrong or undesirable; your job is to rate just exactly how undesirable. Be careful not to rate behaviors which differ in wrongness or undesirability on the same place on the scale. Remember, the more wrong or undesirable the behavior, the further to your right the block should go.

"One last thing—the blocks must be removed from the slot when I read you the items. I will tell you when to remove them.

"Feel free to ask questions about the procedure, or the meaning of any word or phrase, at any time. Do you have any questions now?"

All pertinent questions were answered, then the rating board items were presented to each pair of subjects: first the 30 Mubs, second the Mubs embedded in Vubs (experimental sample) or Amubs (control sample). After completing these tasks jointly, the subjects were taken to separate experimental rooms, where each was told: "I have some more items for you to rate. Instructions are on the front page of the booklet. After you finish, please fill out this short questionnaire."

The questionnaire was designed to tap attitudes and feelings of subjects toward their partners and the experiment, as well as to check that subjects did not know each other.

The sequence of items rated by the separate subjects was identical with the one they had rated during the second rating board presentation: Mubs embedded in Vubs for the experimental sample, and Mubs embedded in Amubs for the control sample.

#### *Unusable Subjects*

Demonstration of norm shift was contingent upon subjects discriminating Mubs from Vubs items and rating different items at different points on

the scale. Pretests showed that a few subjects rated many Mubs items near the "most undesirable" end of the scale; apparently they viewed behaviors either as desirable or undesirable, with few or no gradations in between. Therefore a decision was made, before the main experiment, to set aside data from subjects with this two-valued orientation. Data from subjects who rated ten or more Mubs items 50 or below were to be set aside. Six pairs of the experimental sample and four pairs of the control sample met this criterion for rejection. Additional subject pairs were run to obtain 25 usable pairs for both experimental and control samples.

# RESULTS

## *Subhypothesis 1: Norm Formation*

To demonstrate that a group norm was actually formed by partners, the closeness of partners' ratings was compared with the closeness of nonpartners' ratings. If the between-partner variance is significantly less than the item variance, then the ratings of the two partners, on a given item, are closer together, on the average, than the ratings of two nonpartners. The significance of the differences between these variances may be determined by use of the *F* test. The results of this analysis are shown in Table 1.

From the significance of the *F*-ratios in Table 1 we can see that the ratings of a given item by partners were significantly closer together than ratings by nonpartners. We can therefore infer that a group norm is operating. It should be noted that partners adhered to this norm even during presentation 3, when they made their ratings in different rooms.

## *Subhypothesis 2: Differentiation of Moderately Undesirable Behavior Statements from Very Undesirable Behavior Statements*

To demonstrate that subjects actually did judge the very undesirable behavior statements (Vubs) more undesirable than the moderately unde-

TABLE 1  
*Comparison of Between-Partner and Intraitem Variances for the Moderately Undesirable Behavior Items*

Sample	Presentation	Between-partners Variance	Intraitem Variance	F
Experimental	1	114.87	3484.92	33.1*
Experimental	2	102.13	3479.97	34.0*
Experimental	3	1827.74	4258.01	2.33*
Control	1	120.08	3515.27	29.3*
Control	2	115.34	3767.00	32.7*
Control	3	2016.57	4502.23	2.24*

\* Significant at the .001 level.

sirable behavior statements (Mubs) and the latter about as undesirable as the auxiliary moderately undesirable behavior statements (Amubs), the differences in item ratings between Mubs and Vubs and between Mubs and Amubs were tested.

The Mubs were indeed rated as less undesirable (mean 142.54) than the Vubs (mean 13.64). There is no overlap whatsoever between the item means of the two series; the most undesirable of the Mubs was rated 31.38 millimeters less undesirable than the least undesirable of the Vubs. A *t*-test, confirming this rather compelling difference, showed the mean Mubs rating was significantly higher than the mean Vubs rating at the 1 per cent level. This difference would be even larger if ratings of the Mubs items during the first presentation were omitted.

The Amubs mean of 147.28 was quite close to the Mubs mean 142.54; it is just 4.74 millimeters higher (less undesirable). A *t*-test indicated that this difference was not significant at the 5 per cent level.

From these analyses we may conclude that the classification of items into Mubs, Vubs, and Amubs was congruent with the subjects' actual ratings of the items.

### *Subhypothesis 3: Change in Norms*

To demonstrate a change in norms, it was necessary to show a shift in ratings. According to our major hypothesis, the ratings of the Mubs by the experimental sample should shift in the contrast (less undesirable) direction when the Mubs are embedded in Vubs. The control sample, for whom the Mubs were embedded in Amubs, should show no such shift, since for them judgmental anchorages were not changed.

Table 2 presents the mean ratings of the Mubs items for the experimental and control samples for each of the three presentations. To find out whether, for a given sample, there was a significant change in mean ratings from one presentation to another, the magnitude of the mean change must be compared with its standard error.

A summary of the differences in mean rating of Mubs items from one

TABLE 2  
*Mean Ratings of Moderately Undesirable Behavior Items by Presentation and Sample*

Sample	Presentation		
	1	2	3
Experimental .....	137.81	146.23	152.33
Control .....	138.77	137.87	141.23
Total .....	138.29	142.05	146.78



TABLE 3

*Changes in Mean Ratings of Moderately Undesirable Behavior Statements from One Presentation to Another*

Sample	Change from	Mean Change	S.E. of Mean Change	Critical Ratio
Experimental	1-2	+8.42	0.42	20.0***
Experimental	1-3	+14.52	1.33	10.9***
Experimental	2-3	+6.10	1.21	5.04***
Control	1-2	-0.90	0.38	2.36*
Control	1-3	+2.46	1.37	1.80
Control	2-3	+3.36	1.29	2.60**

\* Significant at the .05 level.

\*\* Significant at the .01 level

\*\*\* Significant at the .001 level

presentation to another, for each sample, along with their standard errors is given in Table 3.

The data presented in Tables 2 and 3 show, first, that the experimental and control samples are well matched, for their average ratings on presentation 1, identical for both, differed by less than a millimeter. Contrast effect was exhibited by the experimental sample; they shifted more than 8 millimeters from presentation 1 to presentation 2, while the control sample shifted very slightly in the other direction. This difference between experimental and control samples persisted in presentation 3, the "alone" situation; ratings of the two groups differed by 11 millimeters.

Thus the formation of a group norm (agreed ratings), and its shift upon a change in the stimulus situation has been demonstrated in a prototype laboratory situation.

#### DISCUSSION

The main thesis of this prototype study, that a group norm will change when the stimulus situation confronting the group changes, has been demonstrated by the shift of ratings of the experimental sample between presentations 1 and 2 (Table 3). The mean shift for the experimental sample was more than 8 millimeters in the direction of "less undesirable," while the shift for the control sample was somewhat less than a millimeter in the opposite direction.

The formation of group norms by subjects in this study is similar to that reported by Sherif (5, 6). These norms regulate experience and behavior even when the individual is not in the physical presence of the social group, but there is usually stricter adherence to these norms when the individual is in the physical presence of the social group. We might expect that the



more a particular group serves as a reference group for an individual, the less he would deviate from its norms when not in its physical presence. This study, designed with other objectives, furnishes no evidence on this point.

There seems to be a pronounced tendency for subjects of both experimental and control samples to rate a given item as less undesirable when alone (presentation 3) than when with his partner (presentation 2). It appears unlikely that this difference was caused by the difference between the paper rating scale and the rating board. It is more likely that the difference is part of the general tendency for sanctions against prohibited behavior formulated by a society to be more severe than the average member of that society believes such sanctions should be. A striking example of this is provided by Kinsey *et al.* (1), who show that a large majority of American males are guilty of sex offenses punishable by law.

There was little awareness on the part of subjects in the experimental sample that there was a consistent shift in these ratings from presentation 1 to presentation 2. Typical answers by experimental subjects to the question, "Do you think that any of your ratings changed from the first presentation of an item to the second presentation of the same item?" were: "Yes, both ways," "Unknown," "No," "Yes. Either direction. Didn't remember exact rating before."

Had subjects known their ratings exactly, the shift of ratings would have been very much reduced. During a pretest, subjects made their ratings of undesirability on a numerical scale. Once they attached a number to a behavior statement, they were very loath to change it.

We might infer from this study that social norms will change when there is a change in the stimulus situation to which the norms are relevant. This change may be facilitated by an awareness of the changes in the situation and retarded by the codification, especially in the form of language, of these norms.

#### SUMMARY

Much work has been done in the disciplines of sociology, anthropology, political science, history, and economics on the subject of social change. The present study attempts to demonstrate a laboratory prototype or paradigm of social change, testing the hypothesis that social change, defined in terms of a change in group norms, would result from a change in the stimulus situation in which these judgmental norms are embedded.

Thirty statements of moderately undesirable behaviors, such as "fishing without a license," were presented, with instructions to rate on graphic rating scales as to undesirability, to 50 pairs of subjects, half experimental, half control, three separate times. The first presentation was without embedding material. The second and third were with 14 embedding items

which were rated along with the 30 moderately undesirable behavior statements. For the experimental sample the embedding items were statements of very undesirable behavior, such as "kidnapping a baby for ransom"; for the control sample, they were of moderately undesirable behaviors, similar to the original 30.

For the first two presentations of the 30 items, which took place in a single continuous session, the pair, sitting side by side, were instructed to discuss the undesirability of each behavior with a view toward making similar ratings of the item on a rating board with separate blocks in two parallel 300-millimeter slots, one block and slot for each subject of the pair. For the third presentation, which took place immediately after this group session, the subjects were alone in separate rooms and rated the items on a 300-millimeter paper graphic rating scale.

Analysis of the ratings of the 30 items showed that a group norm was formed during the group session which caused members of a pair to rate a given item similarly even when in separate rooms. These group norms showed a shift, for the experimental sample, in the direction of rating the 30 moderately undesirable items as less undesirable when judging them in the context of the very undesirable behaviors. The control sample, continuing to judge the 30 items in a context of moderately undesirable items, made no such shift. Thus the hypothesis that a change in the stimulus situation will bring about a change in group norms (in this case agreed-upon standards of judgment) was confirmed.

*Manuscript received: June 28, 1956*

*Revised manuscript received: December 26, 1956*

Edwin Cohen

Educational Research Corporation

10 Craigie Street

Cambridge 38, Massachusetts

#### REFERENCES

1. Kinsey, A. C., W. B. Pomeroy, and C. E. Martin, *Sexual Behavior in the Human Male*, Philadelphia: Saunders, 1948.
2. Long, L., "A Study of the Effect of Preceding Stimuli upon the Judgments of Auditory Intensities," *Archives of Psychology*, 1937, 209.
3. McGarvey, H. R., "Anchoring Effects in the Absolute Judgment of Verbal Materials," *Archives of Psychology*, 1943, 281.
4. Rogers, S., "The Anchoring of Absolute Judgments," *Archives of Psychology*, 1941, 261.
5. Sherif, M., "A Study of Some Social Factors in Perception," *Archives of Psychology*, 1935, 187.
6. Sherif, M., *Psychology of Social Norms*, New York: Harper, 1936.
7. Wever, E. G., and K. E. Zener, "The Method of Absolute Judgment in Psychophysics," *Psychological Review*, 1928, 35, 466-493.

## Notes on "Role Differentiation in Small Decision-Making Groups"

D. K. WHEELER, *University of Western Australia*

A paper by Bales and Slater suggests that in small decision-making groups two status types can be recognized (1, p. 281). These two types, or specialists, are designated as the idea man and the best-liked man, and the suggestion is put forward that differentiation in groups typically produces a *task specialist* who can be thought of as "representing" the task values of group members, and a *social-emotional specialist* who represents "other values and attitudes which tend to a certain degree to be disturbed, de-emphasized, threatened, or repressed by the requirements of the emerging task solution" (1, pp. 297-8). How far can this view be accepted?

The data from which these conclusions arise are for four sessions of each of 14 three-to-six-man groups. The principal type of differentiation is shown in Bales' and Slater's Table 4 to be a "separation of the rankings in likes from the rankings on other measured characteristics." Over all sessions there are more cases in which the best-liked man holds top ranking in only that one characteristic than cases of other isolated prominence. It is shown in Table 5 that in the first session over 50 per cent of the best-liked are voted top on Ideas and over 30 per cent are voted top on Guidance. By the fourth session top position in Ideas is held by the best-liked man in only one twelfth of the cases, and in Guidance it is held in about one quarter of the cases.

On the basis of Tables 4 and 5, the authors conclude, "This tendency for the best-liked man not to be the man chosen as having the best Ideas—a tendency which seems to increase over time—raises the question as to whether men of these two types, high on one characteristic but not on the other, will show congruent differences on overt behavior" (1, p. 279). This seems very much like assuming the conclusion. The argument in this part of the paper seems to run as follows:

1. If a man has top rank on one measure, *but not on any other*, he may be considered a "specialist" in the area measured.
2. Because more men hold top rank on Liking, and only on Liking, than any other variable, therefore to gain top rank on Liking indicates a "specialization" in the "social-emotional" role.
3. Those with top ranking on Ideas are "task specialists."
4. Specialists may differ on their interaction profiles.
5. Composite profiles of top men on Ideas show more interaction in the

"task" area. Composite profiles of best-liked men show more interaction in the "social-emotional" area.

6. There are "task specialists" and "emotional-social specialists." After consideration of the data, this writer finds it difficult to agree with the conclusion.

In session 1 (Table 5) the same men held top position on Liking and Ideas in 52.1 per cent of the cases. As there were 14 groups, such a percentage can be accounted for only by ties, of which there must have been a large number. Later statements support this, e.g., "Some subjects do not make any differentiation in their professed likes toward other members. They say they like all members equally, and in most cases strongly" (1, p. 294). In general, it should be possible to get back to the original data from percentage tables. Inability to do so here and at other places throughout the paper considerably handicaps the reader. For instance, it is not possible to tell, in the case of the 42 per cent of best-liked men who failed to differentiate, whether it was 42 per cent of the 14 men (or more, if there were ties) in any one session, or 42 per cent of all best-liked men (however many they were) over all four sessions. Furthermore, unless interviews were held with subjects, it is not possible to know whether those who failed to differentiate did so because "they liked all members equally and in most cases strongly" or because they found it easier to mark many of their fellows 7 or 8 on an 8-point scale. Such questions have important implications for the rankings on best-liked and for the correlations based on such rankings.

No attempt is made to explain why in this first session the percentage of cases in which the best-liked man is voted top on Ideas is so much greater than chance expectation. Attention is drawn to the other sessions, where the position is reversed. In these sessions there are the same 14 groups with the same 64 men. What happened to the 7 or more men (out of the 14 groups) who in the first session were top on both Ideas and Liking? Did they remain top on Ideas, top on Liking, or neither? The same question might be asked about the one third or more of the group who were top on both Liking and Guidance? Should not the investigation of whether there are actual persons who consistently carry these two "specialist" roles start from the first session where apparently more than half the subjects combine them? Here are persons who rank top on both Ideas and Liking, or Liking and Guidance. What changes in their interaction profiles take place as session follows session? If the theory advanced is correct, it might be expected, for those who remain in top position on Liking, but not on Ideas or Guidance, that there would be a decrease in task-oriented interaction and an increase in the social-emotional categories as they took over the "roles" of "social-emotional specialists." But another suggestion may be made—that those qualities (including appearance, manner, personality,

etc.) which cause a man to be voted best liked by the rest of the group may be qualities which are not necessary to problem-solving and, moreover, qualities which are not necessarily indicated by the other measures. Doubt might even be expressed about the reliability of the ratings on best-liked in view of the lack of differentiation. *A priori*, it might, too, be expected that the measures of Ideas and Guidance would be measures of much the same thing. This is confirmed by the statement that in most of the 23 sessions dealt with in Table 6, the top man on Ideas is also top on Guidance (1, p. 279) though it is not consonant with Table 9 (1, p. 289) where it is shown that only slightly over half the time does the same person hold top position on Ideas and Guidance. The writer's factor analyses show the two measures do measure much the same thing.

Bales and Slater say, "the Idea man shows a concentration in the task area, whereas the Liked man shows a concentration in the socio-emotional type of activity" (1, p. 280). This statement is based on composite profiles of initiated and received interaction for the top men on the Idea ranking and the top men on the Liked ranking for the same 23 sessions. Nowhere is it stated how many men there were. There could have been 23 different men, or the profiles could have come from the same 6 men over four sessions.

Certainly the profiles would seem to indicate some differences in task-oriented and social-emotional interaction, but it should be pointed out that there seem to be no differences between the two profiles on categories 6 (Gives Orientation), 7 (Asks Orientation), and 8 (Asks Opinion). If you ask, "Who had the best ideas?" and then analyze only those people who were top on either Ideas or Liking, *but not on both*, it is surely not surprising that those top on Ideas should initiate significantly more suggestions and opinions. Further, category 3 (Agreement) is neglected when testing for differences in social-emotional interaction and category 6 (Gives Orientation) when dealing with attempted answers to problems.

Most significant is that, out of 56 sessions, there are only 23 which can be analyzed to substantiate this conception of "task-specialist" and "social-emotional specialist." It is not stated in how many sessions the best-liked man was also top on Guidance or Ideas, nor how many sessions were eliminated on which ties for top rank occurred, in either Idea or Like ranking. Surely this is an undue selection of the data.

It should also be noted that about 15 or 16 of these 23 sessions were drawn from low-status consensus groups where the agreements of the rankings of members of each other were comparatively low. "Top Man" has a final sound about it, but it is possible that he was only *just* top man; nowhere is there any indication of the extent of the difference between the top man and the next to the top.

It is not the purpose of this paper to quarrel with the concept of quali-

tative differentiation of interaction, but rather to suggest that this idea of differentiation into task and social-emotional areas does not need verification to the extent that it is necessary to postulate persons who take on roles of "task specialist" and "social-emotional specialist."

On the basis of the Bales categories which separate problem-solving attempts from positive and negative reactions, it is not surprising to find differences in interaction profiles between persons or differences in quality of interaction within or between sessions. It does not seem justified, however, on the basis of the 23 out of 56 sessions to suggest that there is a "task specialist" and an "emotional-social specialist." In between 8 and 38 per cent of the cases (it is impossible to be more specific from the data given) the best-liked man is top on Guidance. In terms of the question asked, the rankings on Liking are psychotelic rather than sociotelic and it is doubtful whether such psychotelic choices have (as the factor analyses below indicate) much to do with a problem-solving situation. Moreover, it must be remembered that these 23 cases include as best-liked men at the most only 23 individuals, at the least possibly only 6 (again it is impossible to say from the data given). The number of individuals used is nowhere stated. Individual differences are so great that this seems a very small number on which to build such an elaborate superstructure.

The rest of this present paper is concerned with some analyses of the intercorrelations of Talking, Receiving, Ideas, Guidance, and Liking for both High and Low Consensus groups. Bales and Slater warn their readers to accept these correlations with reserve, but, as they then proceed to make inferences which support their position, it was thought important to test these inferences by factor analysis. This technique enables us, in dealing with an intercorrelation matrix, to replace the original set of  $n$  variables by  $r$  factors (and also gives us information about the measures of these variables). Probably the authors were not thinking of this sort of analysis when they talked of factors in connection with correlations (Table 8) but their conclusions can be tested in this way.

The following statements are made:

"Even in the High groups, Liking tends to split off as a separate factor from the factor which combines the other four characteristics" (1, p. 286).

"In the Low groups the tendency for Liking to split off as a separate factor is accentuated to the extent that it no longer shows any significant correlations" (1, p. 287).

"Thus, in the Low status consensus groups there seems to be some tendency towards three separable factors, [1] Talking and Receiving hang together, but are not highly correlated with [2] Ideas and Guidance. [2] Ideas and Guidance tend to hang together, but are not significantly correlated with [3] Liking. [3] Liking, in turn, is not significantly related with [1] Talk-



ing and Receiving. This is the picture we were led to expect by the line of reasoning above" (1, pp. 288-289).

This, however, is *not* the picture that emerges when a factor analysis is made. Both correlation matrices for High and Low consensus groups were factorized, using Burt's simple summation method with iteration. For the High status consensus group, one factor only can be extracted, giving the following factor loadings:

Talking	.91
Receiving	.91
Ideas	.88
Guidance	.86
Liking	.56

The largest residual correlation left after the extraction of this single factor is .07 between Talking and Receiving. In terms of this analysis, it is possible to say that there is one and only one hypothetical construct in terms of which the intercorrelations of the various measures may be described. In view of the nature of these intercorrelated variables, it is suggested that this construct is problem solving or task-oriented activity (or, to use Couch and Carter's term, "group goal facilitation"). The slightly higher saturations of Talking and Receiving can probably be explained in terms of the greater objectivity (reliability) of these two measures (due to use of a standardized series of categories).

While it is realized that the intercorrelations for the Low Consensus groups are smaller in size and that those of Liking with the other variables are not significant, nevertheless there seemed sufficient indication of the presence of the same type of single factor to warrant an analysis. Again the intercorrelations can be explained by a single factor as shown:

Talking	.66
Receiving	.71
Ideas	.71
Guidance	.84
Liking	.15

This time, however, the reconstructed correlation matrix is not such a good fit, there being residual correlation of .226 between Receiving and Talking, and .172 between Ideas and Guidance. The first of these might have been expected from prior knowledge of the high correlation between Initiated and Received Interaction, while it is doubtful if the second is significant. Both by simple summation and group factor techniques, it is mathematically possible to extract more than one factor, but the residuals used to obtain it are very small, and it does not seem to be significant. In a



small matrix like this, the use of communalities designed to extract a second factor adds considerably to the factor saturations.

As the intercorrelations for the whole group of subjects were not given, the correlations were averaged for low and high status consensus groups, using Fisher's Z-transformation to give the following table:

	Talking	Receiving	Ideas	Guidance	Factor
Talking					.82
Receiving	.821				.83
Ideas	.654	.614			.82
Guidance	.613	.625	.796		.83
Liking	.290	.345	.318	.368	.40

This, too, can be analyzed into one factor as shown. The largest residual is .134 between Talking and Receiving with .111 between Ideas and Guidance. If an attempt is made to extract a second factor, it is even more clear that it would merely be a statistical artefact contributed by the specific correlations between Talking-Receiving and Ideas-Guidance.

The averaged correlations give an over-all picture, and probably a more reliable one than either of the two groups taken singly. Unfortunately the correlations of the other variables with Leadership were not given; otherwise it would have been possible to use them in these factor analyses and determine whether the single hypothetical construct which seems fairly adequately to explain the intercorrelations had any relation to Leadership.

Considering the three matrices, it might be possible to say that where groups interact to solve a problem, the five ranking measures used have moderately high to high correlations with a hypothetical construct concerned with task activity. Rankings of persons in the group as to degree of liking are related to some extent to this but are presumably also dependent on such things as appearance and the positive or negative emotional tonings of his contributions in the task area. Where consensus about ranking on Guidance and Ideas is low, Guidance rankings seem most highly correlated with this factor of task activity, while Liking rankings seem even more dependent on qualities of interaction not measured by any of the other variables. It would be interesting to know, as these were problem-solving groups, what were the relations between the various groups and the quality of the solutions offered. There is much here of process, but nothing of product.

It is recognized, of course, that with only one variable out of five measuring a relationship (liking another) which may not be directed toward the task, the resulting matrix may be ill-balanced: this might account for the above results. In general, however, the results of these factor analyses do not seem to support the suggestions put forward by Bales and Slater.

## SUMMARY

1. Certain objections are made to the authors' selection and treatment of their data, which, it is held, do not necessarily lead to their conclusions.
2. Factor analyses of their correlational data do not seem to support their argument.

*Manuscript received: September 20, 1956*

*Revised manuscript received: November 13, 1956*

*D. K. Wheeler*

*University of Western Australia*

*Nedlands, Western Australia*

## REFERENCE

1. Bales, R. E., and P. E. Slater, "Role Differentiation in Small Decision-Making Groups," in T. Parsons and R. F. Bales, *Family, Socialization, and Interaction Process*, Glencoe, Illinois: The Free Press, 1955.

## Notes on "Role Differentiation in Small Decision-Making Groups": Reply to Dr. Wheeler

ROBERT F. BALES, *Harvard University*

PHILIP E. SLATER, *Harvard University*

Dr. Wheeler's critique seems to consist of two major objections and a series of lesser ones. Most of the lesser criticisms will be found to trace to the fact that we based our analysis upon sessions as the unit rather than upon individuals. We agree that this mode of treatment leaves many questions unanswered. Dr. Wheeler questions, for example: how many separate individuals were involved in the 23 paired profiles of best-idea and best-liked men? To provide this information: the 23 Idea profiles were produced by 17 separate men, and the 23 Liked profiles by 16 separate men. We agree with Dr. Wheeler that it would be desirable to have a careful study of the Idea and Like rankings of individual men as they change over the four sessions, along with the changes in interaction profiles of the men. But the desirability of additional studies should not lead to a derogation of the information in the present one.

The two major objections are not entirely consistent with each other. Dr. Wheeler seems to be saying that:

- a. The data do not show that Like rankings are significantly different from Idea rankings.
- b. There may be significant differences in interaction between top men on Likes only and top men on Ideas only, but it is inappropriate to refer to such persons as "specialists."

With regard to (a), there is an attempt to cast doubt on the reality of differentiation between Like ratings and Idea ratings by the following arguments: [1] that, without interviewing, one does not know what subjects mean when they rate on the Liking scale, [2] that because a considerable proportion of best-liked men do not differentiate in their likings of *others*, the ratings that others make of *them* are unreliable, [3] that the criterion of "top man" on a given variable is unreliable because one may be just *barely* top, [4] that the Like ratings may be different from the Idea ratings after all, but are irrelevant, since "the rankings on Liking are psychotelic rather than sociotelic and it is doubtful whether such psychotelic choices have much to do with a problem-solving situation," [5] that one may doubt whether enough individuals contributed to the finding of differences to make it worth while to "build such an elaborate superstructure," and [6] that a factor analysis based on a matrix in which only *one* "social emotional" var-

iable is included does not show liking highly enough correlated with any other variable to reveal a "social emotional factor."

With regard to the first four arguments: when a series of statistically significant findings are obtained with a set of measures, as they were in this case, one may wish he had still more reliable measures, but he ordinarily does not contend that the findings are not significant after all because the measures might be too unreliable to produce significant findings in the first place.

As to argument [5]: Dr. Wheeler objects to using only 23 sessions out of 56 as "undue selection of the data." We have reported the actual number of individuals involved above, and it is respectable. However, there seems to be a basic misunderstanding of what we were trying to do. The problem of assessing the frequency of occurrence of a given degree of specialization is quite different from the problem of describing its characteristic quality when it does occur. Our selection of "pure cases" (top on one criterion, not on the other, exclusion of ties) is for the purpose of studying qualitative differences, not, as Dr. Wheeler seems to assume, for the purpose of proving that differentiation always occurs.

There is another method of studying the qualitative differences which can be applied over the whole population of subjects, rather than "pure cases," but it requires groups of equal size. This method consists in comparing the profiles of all men ranked higher on Ideas than they are on Liking, with all men ranked higher on Liking than they are on Ideas. Such a study was carried out by Mann on a subsequent sample, as cited in our paper (1). The findings are essentially similar.

We do not say, as Dr. Wheeler states, that "there is a 'task specialist' and an 'emotional-social specialist'" (*italics ours*), as if a definite and high degree of specialization by person occurred under all conditions. Dr. Wheeler seems to feel that we assert that this division of labor by persons *always* occurs, and it disturbs him because our data show that it occurs (with no ties) under our conditions in less than half the sessions (ignoring the differences between first and later sessions). We have never maintained that a high degree of specialization appears under all conditions, nor (to deal with an unspoken but perhaps implicit argument), that such specialization is a desirable aspect of group process. Our data indicate that it occurs (in the degree specified *less* frequently than one would expect (if the two ratings were associated only by chance) in the first session, and *more* frequently than one would expect in the later sessions. In all, there were 11 sessions in which the same individual held top ranking on both measures, as compared to 23 sessions in which two different persons held top ranking. The 22 remaining sessions showed a tie in one of the two rankings.

With regard to [6], the factor analysis: we know of no way to extract a

common factor from a matrix in which the supposed factor is represented by only a single variable, as in this case. The fact that in the combined matrix the single factor accounts for about 16 per cent of the variance of the Liking rating is not impressive, for two reasons. First, the use of *z*-transformations on rhos based on such small *n*'s overestimates the correlations in the combined matrix. Second, if more "social emotional" tests had been included there is ample evidence from other studies to suggest that a second factor would emerge, and the Liking loading might well disappear in rotation. We hope to report soon a factor analysis of some 32 variables including self-, peer- and observer-ratings on groups similar to these, which should clarify the issue more quickly than controversy over a factor analysis inappropriately applied. Task ability and likeability constitute separate, orthogonal, factors in this later analysis.

Apparently Dr. Wheeler does not himself take his objections under (a) too seriously, since in other parts of his critique he seems to accept our findings, but takes the position that they are little to be wondered at. He feels that "it is not surprising" to find differentiation each time we seem to have found it and wonders that we were not surprised to find such a lack of it in our first sessions. Since we view the habit of shoulder-shrugging at social science findings as the prerogative of unsympathetic laymen, we are willing to be surprised if he will. We began with an expectation of nondifferentiation much like his, and which the structure of first sessions seemed to support. Since that time, so many studies have appeared or come to our attention which show likeability and task ability to be independent variables over a series of conditions that we are quite willing to admit our error in taking this first-session phenomenon for granted.

Dr. Wheeler's second major objection (b) is to the use of the term "specialist" as applied to our top men. The objection is based in part upon the assumption that the use of the term requires continuity of top position through a series of sessions. This seems to us to be a matter of arbitrary choice. We feel that successful or even consistent specialization over a long period is not necessarily prerequisite to the use of the term, and that to speak of specialization within a single session does no violence to common sense. The argument of our paper was that specialization is something that in our groups develops from small beginnings. Where in the process of development one should start using the term is something nature does not tell us.

In summary, there are two points in Dr. Wheeler's attack which we feel are well taken. One revolves around the fact that in some respects our data were not ideally suited to our problem, and were, in fact, gathered for other purposes. We have attempted to cope with this issue through additional studies (2, 3, 4). The second point concerns the consistency or variation over

time of "specialized" behavior, a hiatus which we hope will be closed in future research.

Robert F. Bales  
Laboratory of Social Relations  
Harvard University  
Cambridge 38, Massachusetts

## REFERENCES

1. Mann, R. D., "The Relation of Informal Status to Role Behavior in Small Discussion Groups," unpublished honors thesis, Harvard College, 1954.
2. Slater, P. E., "Role Differentiation in Small Groups," *American Sociological Review*, 1955, 20, 300-310.
3. Bales, R. F., "Task Status and Likeability as a Function of Talking and Listening in Decision-making Groups," in L. D. White (ed.), *The State of the Social Sciences*, Chicago: University of Chicago Press, 1956.
4. Borgatta, E. F., and R. F. Bales, "Sociometric Status Patterns and Characteristics of Interaction," *Journal of Social Psychology*, 1956, 43, 289-297.



## A Note on "Differential Mediation of Social Perception as a Correlate of Social Adjustment"

DANIEL GLASER, *University of Illinois*

The theme of this note is that subcultural differences are more plausible than the conception of psychopathy presented by Baker and Sarbin as an explanation for their findings on a comparison of delinquents with nondelinquents in the performance of a social perception task (1).

### SYNOPSIS OF THEIR COMPARISON

In the study cited a social perception task was administered individually to 48 urban high-school students with no known record of delinquency and to 41 institutionalized delinquents with three or more convictions for serious offenses whose "records seemed to indicate that their transgressions arose from the impulsive gratification of immediate needs rather than from submission to peer group pressures or as the responses of a neurotic seeking punishment." The task consisted of predicting the choices made by a photographed youth in each of two similar movies. Each movie opens with a youth and a psychologist discussing a film about flying. The psychologist then instructs the youth on choosing the activity he would most enjoy doing himself and the activity he would least enjoy in each of ten sets of three magazine pictures of men in various recreational and occupational pursuits. The movies terminate as the photographed youth is about to make his selections. After each movie the subjects in the social perception test were shown the ten sets of magazine pictures and were asked to predict the choices made by the youth in the movie. They were advised of the correctness of each pair of predictions before being shown another set of three pictures. The subjects were motivated by advice that they were helping to test a potential Air Force pilot selection device, and that pilots often must know how to size up new men rapidly.

The movies were made especially for this study, utilizing hidden sound cameras, so that the filmed youth did not know that he was being photographed during his discussion with the psychologist. A nondelinquent youth was utilized for the first film and a delinquent for the second, although this, of course, is not indicated in the movies.

The nondelinquent subjects were somewhat more accurate than the delinquents in predicting the choices made by the photographed nondelinquent youth but slightly less accurate in predicting the choices made by the filmed delinquent youth. These differences were not statistically significant

for this sample. For both movies, however, the nondelinquent subjects improved in accuracy more than the delinquent subjects, improvement being measured by comparing accuracy on the first five predictions with accuracy on the second five. This finding is interpreted as indicating that the delinquents are less able to "profit from experience." Furthermore, the delinquent subjects made more similar predictions for the two filmed youths than did the nondelinquent subjects. This is interpreted as indicating that nondelinquents have greater powers of social perception than delinquents, perceiving others in terms of "differential role concepts," while delinquents tend "to assimilate the roles of others into a single, undifferentiated cognitive structure." Such an interpretation supports a theoretical conception of psychopathy as resulting from retarded development of role-taking or social perception ability, notably the Gough formulation (3), which the authors cite.

#### ALTERNATIVE INTERPRETATION

An alternative explanation for the findings described above suggests itself if one is familiar with persons like the delinquent subjects studied, who were in correctional institutions after a record of three or more serious offenses. As has been observed by Sutherland (7), Tannenbaum (8), Cohen (2) and others, delinquents who have experienced several years of conflict with the law, as did these thrice-convicted youth, progressively are segregated into a subculture of their own. Each successive offense and subsequent glow of success and flight from arrest, each arrest and subsequent dramatization of their offense in trial, and each of the many months of detention, increase the subjects' conception of themselves as in conflict with the authorities of the conventional world. In their years with this conflict experience, the occupational and other social mobility ideals presented by schools and mass media become more inaccessible and unthought of, and the delinquents instead develop conceptions of their ideal selves (or "ego ideals") as master criminals, underworld "big shots," or racketeers of one variety or another. The lone offenders as well as the "socialized delinquents" receive social support for negative valuation of conventional goals and preoccupation with criminal goals because they all are thrown into intimate contact with each other, on a twenty-four-hour basis, when confined. This is especially true in so-called progressive institutions like most juvenile training schools, which have replaced cells with dormitories and, as a result, offer virtually no privacy to their inmates. Thus, when confined, both lone and group offenders share common interests and problems and are relatively isolated and alienated from the interests and problems of nondelinquents. When not confined and not engaged in further crime, their delinquency record largely limits them to lower-class occupations.

The examples which Baker and Sarbin present in reporting the occupational choices involved in their social perception task are "engineer," "doctor," and "biologist"; the recreational choices reported as examples are "playing chess," "a boy and girl dining in a restaurant," and "boys and girls at a beach party." Surely, such predominantly middle- and upper-class occupational ideals and recreational activities are sufficiently more experienced and contemplated by nondelinquents than by thrice-convicted delinquents to account for what differences in discrimination these two groups manifested in predicting the choices which the photographed youths would make. The reasonable assumption that the alternative choices are between activities relatively unfamiliar to delinquents should lead one to expect delinquents to be less accurate than nondelinquents in predicting choices between these alternatives, slower in improving their predictions, and less discriminating in differentiating the predictions of the photographed delinquent from the predictions of the photographed nondelinquent. Also suggesting that delinquents do not have as definite attitudes toward and conceptions of the alternative activities as do nondelinquents are the facts that both the nondelinquent sample and a pilot sample of college and high-school youths were markedly more accurate in predicting the choices of the photographed nondelinquent youth than in predicting the choices of the photographed delinquent youth, while the delinquent subjects had almost identical low accuracy scores for both photographed youths.

An additional consideration which may account for some of the differences in the performance of the two groups studied is the difference in their motivation. It is well known by those working in correctional institutions that inmates frequently are hostile toward persons representing the outside world, or at least ill at ease with them, and are poorly motivated toward performance of tasks assigned by such persons. Consequently, inmates frequently operate at much less than their maximum ability when performing on tests, such as intelligence tests or this social perception test.

In introducing their research, Baker and Sarbin identify psychopathy with "the class of delinquents" which they tried to sample, and describe the psychopath as "an individual who is retarded or suffers a deficiency in the ability to perceive the role-of-the-other and is consequently unable adequately to predict the behavior of others." They assert that this observation is supported by numerous case and analytic studies. It is our impression, however, that most such studies report the existence of role-taking deficiencies when delinquent subjects fail to show appreciation of what the researcher considers the "good intentions" of others, yet at the same time, these researchers report the impressive skill with which these delinquents play "confidence game" tricks requiring sensitivity to and manipulation of the attitudes of others. Thus Redl and Wineman (6) (whom Baker and Sarbin cite), in their chapter on "The Ego That Cannot Perform," point

out that in the treatment situation their emotionally disturbed delinquents have difficulty in "assessing social reality," for they do not see the expectations of the treatment group; when they learn the utility of a "social tool" like giving things away, they do not know its limits. In the next chapter, however, Redl and Wineman report the "delinquent ego" as engaging in "Mechanized Warfare with Change Agents," and they describe their delinquents as having "diagnostic acuity in battle relevant areas" because they are so sensitive to whether or not adults are dangerous or safe to their delinquency, sensing this from subtle shifts in the tone of voice of an adult, the effects of fatigue on the adult, and so forth.

Actually, the life of the thrice-convicted offender promotes development of a high degree of role-taking ability in connection with the positions and roles which such a delinquent encounters, which are different from those encountered by nondelinquents and are not represented on such a social perception task as that reported by Baker and Sarbin. The continuous insecurity and diverse social pressures involved in living in a world of delinquents but getting along with guards and other nondelinquents requires that the delinquent "be all things to all men" and "live by his wits." As Ohlin (4), Oman (5), and others have indicated, inmates react differently to what they recognize as different types of delinquents and to different types of nondelinquent personages. In Turner's terms (9), adjustment to the life of a repeatedly convicted delinquent involves learning to take the role-of-the-other *reflexively*, but without *identification* with the other.

The term "psychopath" has been criticized for its vague and inconsistent reference in most usage, as well as for its connotations that the behavior to which it is applied is innately determined and warrants no further etiological analysis (7). The portrayal of the psychopath as a man "without conscience" suggested the possibility of defining the psychopath as a person with deficient role-taking ability, a definition which could be operationalized by role-taking tests. The confidence man, however, epitomizes both lack of conscience and high role-taking ability. One might suggest, therefore, that the persons most frequently denoted as psychopaths could be subsumed more comprehensively and exclusively if we defined a psychopath as a person lacking normal habits of developing and maintaining identification with others. A major pitfall in devising a test for this deficiency, however, may be that when dealing with persons having strong identifications with members of an in-group whose subculture emphasizes suppression of identification with members of out-groups, one would test only for identification with members of out-groups. This in-group out-group identification differential is connoted by such expressions among criminals as "do your own time" and "never give a sucker a break," and by such expressions in more legitimate entrepreneurial circles as "business is business."

If one should choose arbitrarily to define psychopathy by the single cri-

terion of an over-all deficiency in role-taking ability, then a relatively culture-free test should be sought. Such a test, if possible, would present roles and role-takers which could be assumed to be equally familiar to all subjects tested. It would then, presumably, demonstrate whether delinquents and nondelinquents differ in this ability. The writer would hypothesize that delinquents are highly diverse in this ability, as they are in virtually every other ability for which they have been tested, none of which have markedly differentiated delinquents from nondelinquents. The writer would further hypothesize, however, that role-taking ability varies directly with success in "getting away with" delinquent acts, as opposed to being apprehended and punished. If this hypothesis is valid, one would expect institutionalized delinquents to manifest somewhat less role-taking ability than youth of similar delinquency experience who are not institutionalized.

*Manuscript received: August 27, 1956*

*Revised manuscript received: October 30, 1956*

Daniel Glaser

Department of Sociology and Anthropology

University of Illinois

Urbana, Illinois

#### REFERENCES

1. Baker, B. O., and T. R. Sarbin, "Differential Mediation of Social Perception as a Correlate of Social Adjustment," *Sociometry*, 1956, 19, 69-83.
2. Cohen, A. K., *Delinquent Boys*, Glencoe, Illinois: Free Press, 1955.
3. Gough, H. G., "A Sociological Theory of Psychopathy," *American Journal of Sociology*, 1948, 53, 359-366.
4. Ohlin, L. E., *Sociology and the Field of Corrections*, New York: Russell Sage Foundation, 1956.
5. Oman, M., "The Delinquent in the Training School—An Analysis of Interpersonal Relations and Discipline," M. A. dissertation, University of Illinois, 1952.
6. Redl, F., and D. Wineman, *Children Who Hate*, Glencoe, Illinois: Free Press, 1951.
7. Sutherland, E. H., and D. R. Cressey, *Principles of Criminology* (5th ed.), Chicago: Lippincott, 1955.
8. Tannenbaum, F., *Crime and the Community*, Boston: Ginn, 1938.
9. Turner, R. H., "Role-taking, Role Standpoint, and Reference Group Behavior," *American Journal of Sociology*, 1956, 61, 316-328.

## Psychological Predisposition and/or Subcultural Participation: Reply to Dr. Glaser

THEODORE R. SARBIN, *University of California*

BELA O. BAKER, *Los Angeles State College*

Dr. Glaser's critical note calls us to task on three counts: [1] He argues that the reported differences in mediation of social perception are artifacts of our method. Our illustrations of the preference test suggest a sampling of behaviors not representative of the experiences of the delinquent subjects. This apparent lack of representativeness places the delinquent subjects at a disadvantage. [2] The obtained differences, Dr. Glaser suggests, may in part be explained by differences in task motivation in taking the social perception test. [3] Dr. Glaser asserts that our theoretical orientation—that psychopathic delinquency is associated with perceptual cognitive arrestment and its sequelae—is inadequate as an explanation of our findings and of clinical observations made by others. Instead, participation in the delinquent subculture is invoked as an explanatory principle.

Our reply to Dr. Glaser's first point is simple: Both for the delinquent object and the nondelinquent object, delinquents made as many accurate predictions of first preferences as did the nondelinquents. If differential familiarity with the occupations and recreations represented in our preference test had been a contributing variable, then the predictive accuracy of the delinquents would have suffered. Neither delinquent subjects nor objects had more difficulty in identifying the test items than nondelinquent subjects or objects. Furthermore, the content of the preference test was not so biased as one might have inferred from the illustrations in our article. Here are some additional illustrations: one item showed a truck driver backing a large truck and trailer, a mechanic working on a car, and a service station attendant; another item showed a laborer working on a drill rig, a farm hand pitching hay, and a gardener planting; still another item showed a workingman driving a tractor, a musician playing a bass fiddle, and a factory worker repairing a refrigeration unit.

Dr. Glaser's second point is, of course, applicable to all experimental work with humans where control groups are used. In setting up an experiment, one always asks the question: how can we equate the motivations of the groups? In the beginning, we designed our procedures and our cover story to capture the imagination of adolescent boys. Included in our experiment, for example, was a miniature anti-aircraft gun which the subjects fired at aerial targets. Another task involved dart-throwing at preselected



targets. No potentially boring verbal procedures such as inventories or intelligence tests were introduced into this experiment.

The only way we could know if our procedures motivated the groups equally was by indirect assessment of interest, effort expended, and ego-involvement. All the subjects were rated by five experimenters during their performances on the various experimental procedures, including the social perception test described in our paper. The ratings had to do with the expression of interest in the task, with the amount of effort expended in seeking solutions, etc. Although expected individual differences were found, no group differences in rated task motivation turned up. In addition, a tally of the "number of guesses" on the Street-Gestalt test cards not correctly identified showed no differences between our groups. This aspect of our study has been reported elsewhere (1). Both groups appeared to expend the same effort in finding solutions.

Another indirect assessment of interest in the experimental procedures is the number and kind of impromptu "comments" made by the subjects. Contrary to popular opinion, delinquent subjects do not make more "negativistic" comments than nondelinquents. In fact, on one procedure, the delinquents offered more spontaneous comments that were classified as "task-oriented."

Dr. Glaser's third point is a proposal that subcultural differences—a variant of the differential association theory—may be invoked as an *alternative* explanation for our findings. With clarity and vigor, he shows how a thrice-convicted delinquent may acquire certain role-taking skills under consistent reinforcement in detention halls and correctional settings. His argument that the delinquent is capable of social learning under certain conditions carries the implication that we hold a contrary position—that delinquents are social aments. This implication does not follow from our theory.

Our main emphasis is on the perceptual-cognitive characteristics of youths which result in their *recruitment* into the delinquent subculture. As the social ecologists have demonstrated repeatedly, not all residents of delinquency neighborhoods come to the attention of law-enforcement officers. Living in a complex and pluralistic culture, the child with adequate perceptual-cognitive abilities can acquire those psychological traits necessary for effective social interaction. Those who fail to acquire such traits (through constitutional and/or acquired arrestment in perceptual-cognitive abilities) cannot tolerate delay, boredom, frustration, tension, or complexity. They act impulsively and, more often than not, are rejected by the agents of the core culture. In delinquency areas, the existing subculture available to such rejectees are usually the neighborhood gangs. Here the arrestment of perceptual cognitive abilities is not a liability—in fact, it may be an asset. The impulsive behavior which brought about the youth's rejection by



TABLE 1  
*Uncorrected Accuracy Scores on Social Perception Test*

		Controls N = 49	Institution A N = 18	Institution B N = 23
Nondelinquent object	Mean	13.80	11.83	11.43
	S. D.	3.41	4.74	2.87
Delinquent object	Mean	10.82	9.72	12.35
	S. D.	1.99	2.53	2.18

agents of the conventional culture may be interpreted by the delinquent subculture as daring, courageous, and risk-taking. By contrast with the demands of the pluralistic conventional culture, the monolithic delinquent subculture places less of a premium on the full development of perceptual-cognitive abilities.

What we are trying to say is that certain psychological characteristics predispose certain youths to conduct which, under appropriate supporting conditions, disqualify him for participation in the conventional culture and qualify him for participation in available subcultures. Further, these predispositions may not militate against his social learning under the specific conditions of consistent reinforcement both in the delinquent subculture and in the subculture of the correctional institution.

Stimulated by Dr. Glaser's critical note, we re-examined our data. Our delinquent subjects were drawn from two institutions. From institution A, a county Detention Hall, we had tested 18 subjects who had had three appearances in court but who had not been committed to the state Youth Authority. In institution B, a maximum security reformatory of the Youth Authority, we had tested 23 subjects who had had at least three convictions before being remanded to the Youth Authority. The subjects in institution A had a somewhat shorter history of delinquency than the inmates of institution B. A cultural analysis would suggest that the latter were more enculturated to the delinquent subculture.

By reanalyzing our data, we find some interesting results. In making predictions about the nondelinquent object, both the "less enculturated" A delinquents and the "more enculturated" B delinquents were undifferentiated from each other, but were less accurate than the controls.<sup>1</sup> For the delinquent object, however, the predictions are different. The mean uncorrected accuracy score of the "less enculturated" delinquent group was less than chance, for the "more enculturated" delinquent group the mean ac-

<sup>1</sup> In the experiment proper, we reported only the number of correct predictions of first choices made. Here we score both first and second choices for each three-choice item of the preference test.

curacy was above chance, and even above the mean scores of the nondelinquent subjects.

These results are suggestive and account for the and/or conjunction in the title. The end result of enculturation and socialization processes that we recognize as the typical delinquent may be explained (at least in part) by invoking a theory of psychological predisposition *and* a theory of sub-cultural group membership.

Through the use of a more refined scoring system, it appears that the non-delinquent object is perceived with greater accuracy by the nondelinquent group. Although the ordering of the means for the delinquent object clearly supports the conjunction of the two theories, results for the non-delinquent object are—for the present—indeterminate as support for either theory.

It would require too much space for a discussion of Dr. Glaser's peripheral reference to the distinction between reflexive role-enactment and identification. This and other fine points—such as the validity of the imputation of organized interpersonal sentiments to incarcerated delinquents—will be discussed more fully in later reports.

Grateful for the stimulation afforded by Dr. Glaser's critical note, we shall present a more detailed reply upon the completion of a study now in progress in our laboratory. In this study, differentiation is more easily assessed because we have increased the number of items in the preference test, the number and variety of social objects, and have given additional attention to differential enculturation of the subjects.

Theodore Sarbin  
Department of Psychology  
University of California  
Berkeley, California

#### REFERENCE

1. Jones, D. S., N. Livson, and T. R. Sarbin, "Perceptual Completion Behavior in Juvenile Delinquents," *Psychological Reports*, 1955, 5, 141-146.

## Group Size: A Determinant of the Quality and Stability of Group Decisions<sup>1</sup>

ROBERT C. ZILLER, *University of Delaware*

Research into the relationships with the variable of group size has generally focused upon two types of problems: (a) the relative performance of groups and individuals (3, 11, 12, 19, 20) and (b) the processes and products of groups of various sizes (1, 2, 3, 6, 7, 8, 16, 17, 18). The literature with regard to individual as compared to group problem-solving performance has recently been summarized (9) and the conclusion drawn that groups are generally superior to individuals. The studies concerning the processes of groups of varying size indicate, with reference to large groups, that there are more frequent contributors (2, 8, 17), there is less opportunity for any one individual to adopt a position of leadership (3), the leader has less influence on the group decision (8), there is less consensus (8), and they are slower on concrete problems and faster on abstract problems (18). The results with regard to the quality of performance are conflicting.

However, none of these studies compares the stability of the performance of groups of varying size or of individuals compared to groups, and only one, (1), involves groups of continuously increasing size, a consideration in the experimental design which permits the use of more powerful statistical techniques. Moreover, group decision-making and group problem-solving are sometimes treated as undifferentiated areas of investigation. In the present experiment, for purposes of sound theoretical development, a distinction is proposed: Group decision-making differs from group problem-solving in that the decision maker must act on the basis of incomplete information, thus precluding a relatively objective evaluation of the proposed course of action.

The experiment presented here explores the relative quality and stability of performance of new groups of varying size in two decision-making situations differing with regard to the number of possible alternatives. The experimental condition involving the number of alternatives was introduced to examine the possibility of interaction between the group size variable and the number of possible alternatives in the decision-making situation in relation to the quality of performance.

<sup>1</sup> Grateful acknowledgement is extended to the members of the Crew Research Laboratory, Air Force Personnel and Training Research Center, Survival Research Field Unit, Stead Air Force Base, Reno, Nevada, who aided in the data collection and to Drs. Leonard Berkowitz, William Haythorn, and Irl Irwin, for their insightful critical reading of the original manuscript.

## EXPERIMENTAL PROCEDURE

*Subjects*

Altogether, 210 subjects were involved in the experiment. The subjects were members of aircrews who had been recently assigned to the B-47 jet-bomber training program. The bomber's manpower complement is three officers, pilot, co-pilot, and observer. In spite of this role differentiation, the members were relatively homogeneous with regard to military rank,<sup>2</sup> training, and background. Most of the subjects were "quadruple rated," that is, each officer had been trained as a pilot, navigator, bombardier, and observer. All planned to make a career in the Air Force, and very few were unmarried. Generally, the subjects were highly experienced officers carefully selected for an important assignment.

*Groups*

This study concerns the decision-making performances of ten individuals and ten new groups ranging in size from groups of two to groups of six members inclusively. New groups of each size were formed by randomly assigning members of B-47 crews to different experimental groups. In no case did any crew member serve as a subject in more than one group. However, since only three to six crews were available during a given hour, in the case of four, five, and six-man groups, it was occasionally necessary to assign two members of a regular crew to the same experimental group.

*Decision-making Situations*

The first decision-making problem described to the group or individuals involved, essentially, a conflict as to whether to abandon a physically unfit and utterly exhausted crew member or to risk capture of the entire group by remaining with him in enemy territory until he was able to evade and escape with the crew. No scoring of this task was attempted. The task merely served to demonstrate how vital group decision-making may become in a survival setting and to minimize possible group development effects on the critical problems to follow.

In the first critical decision-making task, the group or individual was required to decide on the number of dots on a 16-by-21-inch card with 3,159 black dots scattered rather uniformly, yet in no geometric pattern, over a white background. The card was exposed for only 5 seconds. No time limit was imposed on the length of the group discussion.

The second critical decision-making task involved two "education free" judgment problems.<sup>3</sup> In each exercise, a problem was raised and 15 facts

<sup>2</sup> Military insignias were rarely worn by the subjects at the Air Force Base where this study was conducted.

<sup>3</sup> Test 401-B, one of the tests in the Intellectual Talents Battery being developed for use with regard to Air Force Officers and by the Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas.

presented which had a bearing on the solution to the problem. The subjects were instructed to choose 4 of the 15 facts which were most important in reaching a decision as to whether a certain enemy-held island base was being developed as an air base or a submarine base, and what site a manufacturer should choose for a factory. The total task score was the sum of the scores of the two separate items.<sup>4</sup>

In the first critical task, the dot-estimate situation, a multitude of alternatives may be considered. Thus, in an earlier experiment involving essentially the same task (21), individual estimates ranged from 200 to 250,000 dots. In the second critical task, only 15 alternatives were presented, 4 of which were correct. Therefore, an essential difference between the two tasks is the number of alternatives which the group may consider. It should also be noted that the scores on the two tasks are uncorrelated with reference to the small sample of individuals involved in this study (see Table 4).

Following the decision-making session, the group members completed questionnaires pertaining to their satisfaction with the decision-making processes and products. Finally, the answers to the problems were announced and the experiment discussed with the groups in some detail.

#### RESULTS

##### *Quality of Performance*

The results pertaining to performance criteria are presented in Tables 1 and 2. With reference to the first task (dot test) the amount of error decreases rapidly as the size of the decision-making body is increased from one to three men. At this point, however, the size of the error increases, but again diminishes, when the decision-making group is comprised of six men. To test the reliability of this error increment with regard to four and five-man groups, Snedecor's method of regression analyses (15, p. 410) was employed. In this statistical procedure, a significant first-order component indicates a linear trend; second order, a parabolic trend; and a third order, a cubic trend. Thus, with respect to the dot-test results in Table 1, a significant cubic component will establish the reliability of the increment and successive decrement in error of estimate.

Initially, a simple analysis of variance over group size was calculated and found to be statistically significant. In the component or regression analysis, the first three components accounted for the majority of the variance. However, only the linear and cubic components proved to be statistically significant.

On this basis, it may be stated with a reasonable degree of assurance, that error of decision tends to decrease in relation to the increase in group

<sup>4</sup> The correct responses were those agreed upon by a "panel of military and civilian experts."

TABLE 1

*Regression Analysis of the Relative Performance of Individuals and Groups of Different Sizes on the Dot Test*

Group Size		Mean Error Score				
	1	3330				
	2	1081				
	3	850				
	4	1624				
	5	1806				
	6	773				

  

Source	Sum of square	d.f.	Mean squares	F	Required	
					.05	.01
Total	156,518,033	59	—	—	—	—
Within	111,657,978	54	2,067,740	—	—	—
Between	44,860,055	5	8,972,011	4.34	3.34	4.76
a. Linear	13,949,996	1	13,949,996	6.75	4.00	7.08
b. Quadratic	6,782,237	1	6,782,237	3.28	4.00	7.08
c. Cubic	23,996,975	1	23,996,975	11.61	—	7.08
d. Others	130,846	2	—	—	—	—

TABLE 2

*Regression Analysis of the Relative Performance of Individuals and Groups of Different Sizes on the Multiple-choice Task*

Group Size		Mean Score			
	1	3.20			
	2	4.20			
	3	4.85			
	4	5.00			
	5	5.45			
	6	5.75			

  

Source	Sum of squares	d.f.	Mean squares	F	Required at .01
Total	17,624.6	59	—	—	—
Within	13,357.5	54	247.4	—	—
Between	4,267.1	5	853.4	3.45	3.34
a. Linear	3,960.3	1	3960.3	16.01	—
b. Quadratic	220.1	1	220.1	.89	—
c. Cubic	64.2	1	64.2	.26	—
d. Others	22.5	2	—	—	—

size, but that a significant deviation from this linear trend occurs with respect to four- and five-man groups in this task.

The results with regard to the second critical task were also significantly different between the groups of varying size. However, only the linear component proved to be statistically significant. Moreover, the linear component accounted for the majority of variance in this case (see table 2).

In both tasks, then, a linear regression tends to fit the data fairly adequately. Thus, group size and accuracy of group decision appear to be positively related. However, a marked deviation from the linear trend is significant with regard to the first exercise.

### *Member Reactions*

In attempting to explain these results, the questionnaire items are of limited value since they were completed after both decision-making situations had been resolved. Nevertheless, it was hoped that some bit of information might be found to suggest an avenue of inquiry. The results are presented in Table 3.

Only the results concerning the group's satisfaction with their participation (item 4) were statistically significant, and these, contrary to expectations, indicate an increment in satisfaction through four-man groups (see

TABLE 3  
*Questionnaire Responses of Varying Size Groups*

Item No.*	Group Size						Required .05
	2	3	4	5	6	F	
1	8.0†	8.6	8.7	8.4	8.2	—	—
2	7.6	8.7	9.3	8.5	8.8	2.13	2.45
3	7.2	8.7	8.8	7.9	8.5	3.50	2.45
4	7.4	9.0	8.8	8.2	8.1	1.33	2.45
5	9.9	9.6	9.4	9.2	9.2	2.33	2.45

\* Each item was followed by an appropriate 10-point descriptive scale in which the higher score indicated the most favorable response.

† Mean weighted response of 10 groups.

1. All things considered, how satisfied are you with the quality or degree of excellence of the decisions reached by your group?

2. Considering the entire problem-solving session: "My opinion was given adequate consideration."

3. How satisfied are you with your part in this discussion?

4. To what extent do you feel that the participants worked as a unified group rather than a disjointed collection of individuals?

5. Did you feel free to question the suggestions and proposals of the other group members?



TABLE 4  
*Significance of the Difference among Correlation Coefficients of the Scores on Two Tasks  
 for Groups of Varying Size*

Group Size	r	n
1	-.17	10
2	-.03	10
3	-.29	10
4	-.02	10
5	.09	10
6	.71*	10

Chi square = 6.19; d.f. = 5;  $p = .20 - .30$

\*Significant at the .01 level of confidence

Table 3). This pattern is also evident with regard to questions 1 and 3, although in none of these latter cases are the results statistically significant. Thus, in some respects, the four-man groups appear to react most favorably to the decision-making session as a whole.

On the other hand, the results with regard to five-man groups are somewhat in accord with expectations based on the dot-test performance data. With reference again to the item (item 4) concerning participation in the discussion, it is observed that the five-man groups reacted less positively than the three, four, and six-man groups.

#### *Stability of Performance*

To investigate the relative stability of group performance in reference to group size, Pearson correlations were calculated between the pairs of scores for the groups of each size. In order to test the hypothesis that the six samples are from a common population, a chi-square design was employed (4, p. 135). The results are not statistically significant and the null hypothesis must be retained (see Table 4). Furthermore, only the correlation coefficient between scores of the six-man groups is significant.

#### DISCUSSION

##### *The Cubic Effect*

In attempting to interpret the results with regard to quality of performance, a group development effect must be considered since three group decision-making situations were involved (including the introductory task). In this case, the vanishing of the cubic trend in the 15 alternative situations may be attributable to group development, that is, a developing group structure and/or group learning. Assuming that the groups of different sizes have the same potential for effectiveness, we may hypothesize that the four- and five-man groups require a longer "learning" period before their potential is fully realized.

Another possible interpretation of the cubic phenomenon involves both the nature of the task and group size. It may be hypothesized that the need for organization rises sharply in the case of the four-man groups, particularly when a decision-making task with a large number of alternatives is encountered, and that four- and five-man groups often do not recognize the organizational demands of the task. The hypothesized failure to recognize the need for organization under the dot-task conditions may be attributed, in part, to the relatively small size of these groups which may be deceiving with respect to the number of channels of communication opened by the increased amount and dispersion of information or ideas which must be processed (7, 12, 13). However, in a group of six, a threshold of awareness may be reached enabling the members to respond to the now more apparent difficulties of communication by organizing the group in a manner by which messages may be processed more efficiently (10).

Support for the hypotheses that groups of six experience fewer communication difficulties may be found in a study of the relation of group size and interaction profile by Bales and Borgatta (1). Comparisons of the interaction of groups ranging in size from two to six inclusively reveal that the six-man groups were high in showing solidarity and tension release, but lower in showing tension. On the other hand, the four-man groups were higher in disagreement and in demonstrating high tension and antagonism.

Then too, in an experiment conducted by Bass and Norton (3), regarding the emergent group leader in groups with two, four, six, eight, and twelve members, there was a tendency for absolute variation of leadership ratings by observers to be maximum with groups of six. Moreover, maximum agreement was reached by the observers when six participants per group were assessed. These findings may be interpreted as indicating that leadership functions tend to become localized to a greater degree in groups of this size.

### *Stability of Performance*

While the results with regard to stability of performance are not statistically significant, they suggest that the quality of decisions submitted by six-man groups is more consistent than the quality of decisions submitted by small groups or by individuals. Stated somewhat differently, this suggests that a lower probability of consistently high-quality decisions exists for individuals or new groups with less than six members.

### SUMMARY

This study explores the relative quality and stability of the performance of new groups of varying size in two decision-making situations differing with regard to the number of possible alternatives. Fifty groups were involved in the experiment, ranging in size from groups of two to groups of

six members inclusively. The performances of ten individuals were also included in the analysis.

It was established with a reasonable degree of assurance that a positive linear relationship exists between group size and accuracy of group decisions. However, a significant deviation from this linear trend was observed with regard to the task with the larger number of alternatives and with regard to four- and five-man groups. Two interpretations of this finding were offered. A tendency was also observed for a six-man group to perform at a more consistent qualitative level.

*Manuscript received: September 20, 1956*

*Revised manuscript received: December 26, 1956*

Robert C. Ziller  
Fels Group Dynamics Center  
University of Delaware  
Newark, Delaware

#### REFERENCES

1. Bales, R. F., and E. F. Borgatta, "Size of Group as a Factor in the Interaction Profile," in A. P. Hare, E. F. Borgatta, and R. F. Bales (eds.), *Small Groups*, New York: Knopf, 1955.
2. Bales, R. F., and F. L. Strodbeck, "Phases in Group Problem Solving," *Journal of Abnormal and Social Psychology*, 1951, 46, 485-495.
3. Bass, B. M., and F. M. Norton, "Group Size and Leaderless Discussion," *Journal of Applied Psychology*, 1951, 35, 397-401.
4. Edwards, A. L., *Experimental Design in Psychological Research*, New York: Rinehart, 1950.
5. Festinger, L., "Informal Social Communication," in L. Festinger, K. Back, S. Schachter, H. H. Kelley, and J. Thibaut (eds.), *Theory and Experiment in Social Communication*, Ann Arbor, Michigan: Edwards Brothers, 1950.
6. Fox, D., I. Lorge, P. Wertz, and K. Herrold, "Comparison of Decisions Written by Large and Small Groups," *American Psychologist*, 1953, 8, 351 (Abstract).
7. Gibb, J. R., "The Effects of Group Size and of Threat Reduction upon Creativity in a Problem-Solving Situation," *American Psychologist*, 1951, 6, 324 (Abstract).
8. Hare, A. P., "Interaction and Consensus in Different Sized Groups," *American Sociological Review*, 1952, 17, 261-276.
9. Kelley, H. H., and J. W. Thibaut, "Experimental Studies of Group Problem-Solving and Processes," in G. Lindzey (ed.), *Handbook of Social Psychology*, Cambridge, Massachusetts: Addison-Wesley, 1954.
10. Maier, N. R. F., and A. R. Salem, "The Contribution of a Discussion Leader to the Quality of Group Thinking, the Effective Use of Minority Opinions," *Human Relations*, 1952, 5, 277-288.
11. Marquant, D. I., "Group Problem-Solving," *Journal of Social Psychology*, 1956, 41, 103-113.
12. Roby, T. B., and J. T. Lanzetta, "An Investigation of Task Performance as a Function of Certain Aspects of Work-Group Structure," *Research Report*,

AFPTRC-TN-56-74, San Antonio, Texas: Lackland Air Force Base, Air Force Personnel and Training Research Center, June, 1956.

13. Roby, T. B., and J. T. Lanzetta, "A Replication Study of Work-Group Structure and Task Performance" (Manuscript on file, Operations Laboratory, Air Force Personnel and Training Research Center).
14. Shaw, M. E., "A Comparison of Individuals and Small Groups in the Rational Solution of Complex Problems," *American Journal of Psychology*, 1932, 44, 491-504.
15. Snedecor, G. W., *Statistical Methods* (4th ed.), Ames, Iowa: Iowa State College Press, 1946.
16. Smith, E. B., "Some Psychological Aspects of Committee Work," *Journal of Applied Psychology*, 1927, 11, 348-404.
17. Stephan, F. F., and E. G. Mishler, "The Distribution of Participation in Small Groups: an Exponential Approximation," *American Sociological Review*, 1952, 17, 598-608.
18. Taylor, D., and W. L. Faust, "Twenty Questions: Efficiency in Problem-Solving as a Function of Size of Groups," *Journal of Experimental Psychology*, 1952, 44, 360-368.
19. Timmons, W. M., "Can the Product Superiority of Discussion Be Attributed to Averaging of Majority Influence?" *Journal of Social Psychology*, 1942, 15, 23-32.
20. Watson, G. B., "Do Groups Think More Efficiently than Individuals?" *Journal of Abnormal and Social Psychology*, 1928-1929, 23, 328-336.
21. Ziller, R. C., "Scales of Judgment: a Determinant of the Accuracy of Group Decisions," *Human Relations*, 1955, 8, 153-164.



